UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT
DROIT D'AUTEUR

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document. Toute contrefaçon, plagiat ou reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

UNIVERSITÉ DE LORRAINE
34 COURS LÉOPOLD — CS 25233
54052 NANCY CEDEX
TÉL : 03 54 50 54 00
CONTACT@UNIV-LORRAINE.FR
WWW.UNIV-LORRAINE.FR

École doctorale IAEM Lorraine

Influence des dynamiques d'opinion sur le changement climatique : analyse et stratégies macroscopiques

THÈSE

présentée et soutenue publiquement le 8 avril 2026

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention Automatique, Traitement du signal et des images, Génie informatique)

par

Anthony Couthures

Composition du jury

Président :	Romain Postoyan	Directeur de recherche CNRS au CRAN, Nancy, France
Rapporteurs :	Antoine Girard	Directeur de Recherche CNRS au L2S, Saclay, France
	Julien M. Hendrickx	Professeur à EPL & UCLouvain, Louvain-la-Neuve, Belgique
Examinateur :	Federica Garin	Chargée de recherche INRIA Rhône-Alpes, Grenoble, France
Directeur :	Samson Lasaulce	Directeur de Recherche CNRS au CRAN, Nancy, France
Co-Directeur :	Constantin Morărescu	Professeur à l'Université de Lorraine, Nancy, France

Centre de Recherche en Automatique de Nancy - UMR-CNRS 7039
2, avenue de la forêt de Haye 54516 Vandœuvre-lès-Nancy

Remerciements

Étant spinoziste, je ne peux m'empêcher de dire que cette thèse n'est pas uniquement le fruit de ma personne, mais plutôt la résultante de l'ensemble des individus qui m'ont accompagné et poussé dans cette voie. On notera que c'est un peu un retour aux sources de faire cette thèse dans le domaine de la sociologie mathématique.

Mes premiers remerciements s'adressent aux membres du jury pour avoir pris le temps de relire et de commenter ce manuscrit, fait le déplacement pour ma soutenance, et m'avoir challengé avec leurs questions : j'ai adoré l'exercice. Je suis sincèrement heureux de vous avoir eus comme évaluateurs. Je souhaite aussi remercier toutes les personnes qui prendront le temps de lire cette thèse à l'avenir, cela compte énormément.

Je tiens également à remercier le CNRS pour mon financement spécifique, ainsi que pour son extension, sans lesquels je n'aurais pas pu réaliser cette étude, ni prendre le temps de la faire de manière optimale. Enfin, j'ai une pensée pour ce formidable outil qu'est Internet et les plateformes open-source (arXiv, HAL, StackExchange, etc.) sans lesquels la connaissance ne serait pas accessible et la recherche moderne impossible.

Si je suis arrivé jusqu'ici, c'est aussi parce que certaines personnes m'ont mis sur la voie très tôt. Parmi les points de passage les plus marquants, j'aimerais remercier ma professeure de physique du lycée, Mme Lourde Rocheblave, qui m'a compris, m'a pris sous son aile, et m'a évité bien des soucis en me déconseillant de partir en classe préparatoire. Aussi, je remercie Mme Bribet et M. Lourme pour avoir énormément contribué à mon ouverture d'esprit dès le lycée. J'espère qu'un jour vous tomberez sur cette thèse par hasard en cherchant votre nom.

Je remercie infiniment mes encadrants. Tout d'abord, Samson, pour m'avoir permis de découvrir ce laboratoire et tous les gens qui s'y trouvent, pour m'avoir toujours poussé plus loin et amené à avoir une vision globale de la recherche. Constantin, merci d'avoir supporté mes rigidités et d'avoir toujours eu la porte ouverte dès que j'ai eu le moindre problème ou une nouvelle lubie mathématique, tu as su me focaliser dès qu'il y en avait besoin (c'est-à-dire très souvent). Tu es vraiment un chercheur que j'estime énormément et j'espère qu'on aura l'occasion de travailler ensemble à nouveau. Enfin, au-delà de cet encadrement officiel, je tiens à remercier Vineeth, mon véritable "Big Brother" de recherche. Merci d'avoir ajouté du chaos à ma recherche et de subir mes messages le week-end !

J'aimerais aussi remercier les permanents du CRAN avec qui j'ai pu échanger au fil des années : Marc, Mattia, Romain, Pierre, Jamal, et tous les autres. J'ai également une pensée pour mes co-élus du conseil de laboratoire avec qui on a dû gérer quelques galères. Je pense aussi à Olivier Beaude qui m'a invité à faire un séjour à EDF et m'a permis de découvrir le monde de la recherche dans le privé. C'est une expérience qui a été très formatrice (même si elle m'a conforté dans l'idée que je ne veux pas travailler dans le privé).

J'ai une pensée toute particulière pour Christine, qui est véritablement la "maman" des doctorants ici et qui m'a énormément aidé sur le parcours du combattant que peut être l'administratif parfois.

Enfin, je remercie énormément tous les doctorants du CRAN, les anciens comme les nouveaux, pour avoir participé à tout ça. Surtout si vous avez dû me subir à l'heure du déjeuner ou le week-end pendant que je partais dans tous les sens. Je pense à mes cobureaux : Bouchra (ma grande sœur de thèse, avec qui j'ai partagé de nombreuses discussions), Elena (ma cousine de thèse, à qui j'ai volé le bureau) qui m'ont manqué sur la fin, Aurélien, Thomas et Hassan (mes petits frères de thèse, JulJulJul à vous deux), qui ont dû subir les perturbations que je leur ai infligées ! Mais aussi Maxime et Gustave chez qui j'avais une chaise attitrée dans leurs bureaux

respectifs. Plus particulièrement pour vous deux, vous m'avez formé à l'automatique (même si je ne comprends toujours rien au contrôle) avec un point de vue "ingénieur mental" et un point de vue "mathématicien mental" qui m'ont permis de mieux comprendre la théorie à force de vous voir vous disputer sur la façon de m'expliquer les choses. Merci beaucoup d'être vous.

D'ailleurs, au-delà des personnes, je tiens à remercier mon bureau en lui-même (120 jaune). Cet espace que j'aime tant, où je me sens chez moi, et qui a été le refuge de cette recherche.

La ville de Nancy a été la toile de fond de cette thèse avec son climat et son ambiance. J'y ai été accueilli par des colocataires formidables (Antonina, Francesca, Pauline et Giulia). Merci à elles, vous avez été une respiration dans une période de ma vie compliquée. Ensuite, j'ai eu l'occasion d'habiter avec mon ami et rival du supérieur, Adrien, avec qui j'ai commencé à parler de recherche dès notre rencontre et avec qui on s'est encouragés à aller toujours plus loin et donc à faire une thèse (je l'ai fait, tu es le prochain). Merci à mes amis de Bordeaux (à tous les Jean : Robin, Julien, Kilian, Valentin, Camille...), et mes amis de mes études. Je ne pourrai pas tous vous citer (car j'ai peur d'en oublier) mais merci à tous.

Je n'oublie pas l'essentiel : ma famille. J'aimerais tout d'abord remercier mes parents sans qui mes frères et moi ne serions pas à ce niveau d'études. Je ne sais pas comment vous avez fait, mais vous nous avez, à tous les trois (Jeremy, Aurélien et moi), donné le goût de la recherche et de la résolution de problèmes. Je ne suis pas le premier docteur de la famille (à une semaine près, mon frère cadet Jeremy m'est passé devant !) mais c'est tout de même un honneur de porter le nom de Couthures dans le monde de la recherche. Je vous aime tous énormément, même si je ne le dis pas assez souvent.

Ensuite, j'aimerais remercier Claire, mon rayon de soleil et ma future femme, pour me supporter, me canaliser, me motiver et endurer mes intérêts très spécifiques. Grâce à elle, on rigole, on ne s'ennuie jamais, et ça m'a énormément aidé dans ma recherche. Je l'ai rencontrée en faisant visiter la colocation que j'ai mentionnée plus haut, et j'ai été ébloui par son sourire et sa personnalité. Elle m'a accompagné pendant toute la thèse et m'a permis de me sentir moins seul dans cette aventure. Merci d'avoir toléré mon rythme de sommeil décalé et ma fâcheuse tendance à accumuler les réveils inutiles, alors que tu devais te lever tôt pour aller travailler. Je t'aime.

Ce voyage a commencé avec l'adoption de mon chat, Gin, que j'aime énormément et qui m'a accompagné pendant toute la thèse. Celui-ci n'hésite jamais à aller dans sa litière au pire moment de la nuit (et ce, même la veille de la soutenance !) et à me dormir dessus. Merci à lui.

Enfin, pour conclure, je tiens à remercier ce qui a canalisé mon attention pendant ces longues et heureuses heures de travail, guidé ma recherche et permis de trouver toutes ces idées : ma bande-son. Merci aux artistes Daft Punk, Gorillaz, Childish Gambino, Tame Impala, Mac Miller, L'Impératrice, Carbon Based Lifeforms, Stellardrone, Fejka (ainsi que les compilations de la chaîne CureMusic), et bien sûr Lofi Girl qui m'ont permis de faire cette recherche. C'est peut-être peu pour eux, mais ça représente beaucoup pour moi.

I dedicate this thesis to future generations and the hair I lost while writing it.

Contents

Notations **xi**

Chapter 1
Introduction

1.1 From Societal Imperative to Scientific Challenge 1
1.1.1 Why Progress Stalls: Two Key Obstacles 2
1.1.2 From Description to Explanation: The Need for a Mechanistic Understanding 3
1.2 A Journey into Socio-Environmental Dynamics 4
1.3 First Stage: Formalizing Strategic Conflict: A Game-Theoretic Perspective 5
1.3.1 The Initial Question and Toolkit 5
1.3.2 State of the Art: A Disciplinary Fracture Between Two Worlds 5
1.3.3 Our Contributions: A Construction of a Socio-Environmental Game . . . 8
1.3.4 Insight and Pivot to the Next Stage 10
1.4 Second Stage: Interplay of Environment and Opinion Dynamics 11
1.4.1 The Evolving Question and Toolkit 11
1.4.2 State of the Art: The Rise of Coupled Socio-Environmental Models 11
1.4.3 Our Contributions: Uncovering the Mechanisms of Dynamic Behavior . . 13
1.4.4 Insight and Pivot to the Next Stage 15
1.5 Third Stage: The Fundamental Mechanisms of Polarization 16
1.5.1 State of the Art: Models of Opinion Dynamics with community structure 17
1.6 Outline of the Manuscript 20
1.6.1 Contributions 20
1.6.2 Chapter contents 21
1.6.3 List of publications 21
1.6.4 Other activities 22

Chapter 2
Networked Systems Preliminaries **25**

2.1 Introduction . . . 25
2.2 Algebraic Graph Theory . . . 26
2.2.1 Basic Graph Concepts . . . 26
2.2.2 Canonical Graphs . . . 28
2.2.3 Graph Representations: Matrices . . . 29
2.2.4 Graph Partitions . . . 35
2.3 Matrix Eigenvalue Theorems . . . 39
2.4 Dynamical systems . . . 41
2.4.1 Existence and properties of solutions of differential equations . . . 41
2.4.2 Invariant Sets and Asymptotic Behavior . . . 43
2.4.3 Cooperative, monotone, and strongly monotone systems . . . 45
2.4.4 Lyapunov Stability of Autonomous Systems . . . 51
2.4.5 Invariance Principle . . . 54
2.4.6 Input-to-State Stability . . . 55
2.5 Mathematical Tools . . . 55

Chapter 3
A non-linear opinion dynamics model: Continuous Opinion Signaled Action (COSA) **57**

3.1 Introduction . . . 57
3.2 Continuous Opinion Signaled Action (COSA) . . . 57
3.2.1 Model Setup: Agents and Communication Graph . . . 57
3.2.2 From CODA to COSA . . . 59
3.2.3 The COSA model . . . 61
3.2.4 Signal Function and Communication Model . . . 62
3.3 Trajectory Behavior: Invariance and Cooperation . . . 65
3.3.1 Invariance of the state space $\mathcal{X}$ . . . 66
3.3.2 Cooperative Properties and Guaranteed Convergence . . . 66
3.4 Characterization of System Equilibria . . . 69
3.5 The Fully Synchronized (FS) manifold and the FS dynamics . . . 70
3.5.1 Invariance of the Fully Synchronized Manifold (FSM) . . . 71
3.5.2 The Fully Synchronized Dynamics . . . 72
3.5.3 Stability Analysis of Equilibria on the FS Manifold . . . 76
3.6 Preliminary Results on the COSA dynamics . . . 80

3.6.1 Fully synchronized equilibria . . . 80
3.6.2 Invariant Hypercubes Defined by Fixed Points of s . . . 81
3.6.3 Non-Fully Synchronized Equilibria: A First Look . . . 83
3.7 Concluding Remarks and Outlook . . . 85

Chapter 4
The Full Synchronization in COSA **87**

4.1 Introduction . . . 87
4.2 Measuring Disagreement: The Natural Geometry of COSA . . . 87
4.2.1 Motivation for a Weighted Inner Product . . . 88
4.2.2 Weighted Projections and Equivalence of Convergence . . . 88
4.2.3 The Synchronization Error Vector . . . 89
4.3 Stability and Attraction of Fully Synchronized Equilibria (FSE) . . . 90
4.3.1 Local Stability of FSE . . . 91
4.3.2 Characterization of Basin of Attraction for any FSE . . . 93
4.3.3 The case of underestimation and single crossing function . . . 98
4.4 The Interplay of Signal Function and Network Topology . . . 99
4.4.1 Necessary Conditions for the Emergence of Disagreement . . . 100
4.4.2 Synchronization Enforced by Network Topology . . . 102
4.4.3 Disagreement Enabled by Signal Function Nonlinearity . . . 104
4.5 A Sharp Threshold for Global Synchronization . . . 106
4.6 Global exponential stability of the FSM . . . 109
4.7 Concluding Remarks on Full Synchronization . . . 111

Chapter 5
Partial Synchronization in COSA **113**

5.1 Introduction . . . 113
5.2 Perfect Local Synchronization: The Constraint of Structural Symmetry . . . 113
5.2.1 From NFSE to Cell-Synchronized Equilibria (CSE) . . . 114
5.2.2 The Necessity of Structural Symmetry for CSEs . . . 114
5.2.3 A Foundational Example: Twin Node Synchronization . . . 116
5.3 Analysis of Perfect Clusters via the Quotient Graph . . . 118
5.3.1 Model Reduction: The Quotient Dynamics . . . 118
5.3.2 Analysis of the Quotient System . . . 120
5.3.3 Stability of Local Agreement Equilibria . . . 122
5.4 Approximated Local Synchronization: an Input-to-State Stability Concept . . . 129
5.4.1 ISS Analysis of Internal Cluster Disagreement . . . 130

5.4.2 Numerical Illustration: Robust Clustering in the Karate Club Network . . 134
5.5 Cluster Deviation Stability 136
5.6 Concluding Remarks on Local Synchronization 139

Chapter 6
Conclusions and Perspectives **141**

6.1 Conclusions 141
6.1.1 First step of a research journey 141
6.1.2 Manuscript Conclusion 141
6.2 Perspectives 142
6.2.1 First Stage Perspectives: Designing Games for Cooperation 142
6.2.2 Second stage perspectives 143
6.3 Manuscript perspectives 145
6.3.1 Extend the nonlinear interaction model 145
6.3.2 Dynamics of Polarization and Recruitment 148
6.3.3 Community structure and a "new" approach for clustering 149

Appendixes

Appendix A
Paper of the second stage: Interplay of Environment and Opinion Dynamics

A.1 Analysis of a continuous opinion and discrete action dynamics model coupled with an external observation dynamics 152
A.1.1 Introduction 152
A.1.2 Problem Formulation and Preliminaries 152
A.1.3 Analysis of the Model 153
A.1.4 Analysis of the Synchronized Behavior 155
A.1.5 Numerical Simulations 156
A.1.6 Conclusions 157
A.2 Analysis of an opinion dynamics model coupled with an external environmental dynamics 158
A.2.1 Introduction 158
A.2.2 Problem Formulation and Preliminaries 159
A.2.3 Analysis of the general case 160
A.2.4 Analysis of the Fully Synchronized Opinions (FSO) case 162
A.2.5 Numerical results 165
A.2.6 Conclusions 169

A.3 Bifurcation analysis of an opinion dynamics model coupled with an environmental dynamics . . . 173
A.3.1 Introduction . . . 173
A.3.2 Problem Formulation . . . 174
A.3.3 Preliminary results . . . 174
A.3.4 Analysis of Synchronized Agents Dynamics . . . 175
A.3.5 Numerical Simulations . . . 178
A.3.6 Conclusions . . . 178

Appendix B
Paper of the first stage: Formalizing Strategic Conflict

B.1 A static game model for climate change analysis and strategy design . . . 182
B.1.1 Introduction . . . 182
B.1.2 A Game of Emissions and Climate Dynamics . . . 184
B.1.3 Non-Cooperative Equilibrium Analysis . . . 186
B.1.4 Equilibrium Properties: Structure and Uniqueness . . . 188
B.1.5 Inefficiency and the Stability of International Environmental Agreements 191
B.1.6 Numerical Analysis . . . 193
B.1.7 Conclusions . . . 203
B.1.8 Appendix A. The Simple Climate Model (SCM) in Detail . . . 203
B.1.9 Appendix B. Proofs for Section 4 . . . 205
B.1.10 Appendix C. Proofs for Section 5 . . . 211

Appendix C Résumé de la thèse en français

C.1 De l'impératif sociétal au défi scientifique . . . 215
C.2 Un parcours de recherche en trois étapes . . . 215
C.2.1 Interactions stratégiques et théorie des jeux . . . 216
C.2.2 Co-évolution Opinion-Environnement . . . 217
C.3 Troisième étape : Les mécanismes de la polarisation . . . 219
C.3.1 Le modèle COSA et la topologie de l'espace d'état (Chapitre 3) . . . 219
C.3.2 Géométrie naturelle et seuil exact de synchronisation (Chapitre 4) . . . 220
C.3.3 Théorie structurelle des clusters et robustesse (Chapitre 5) . . . 220
C.4 Conclusion générale . . . 221

Bibliography

Notations

Abbreviations

i.e.	*id est*
e.g.	*exempli gratia*
s.t.	*such that*
resp.	*respectively*

Logic

$\forall, \exists$	*for all, there exists*

Set operations

$\times$	Cartesian product
$\cup$	Union
$\cap$	Intersection
$\setminus$	Set difference
$\lvert\mathcal{S}\rvert$	Cardinal of $\mathcal{S}$

Usual sets

$\emptyset$	The empty set
$\mathbb{N}$	The positive integers ($0 \in \mathbb{N}$)
$\mathbb{Z}$	The integers
$\mathbb{R}$	The real numbers
$\mathbb{C}$	The complex numbers
$[a, b)$	$\{x \in \mathbb{R} : a \leq x < b\}$
$[a, b]$	$\{x \in \mathbb{R} : a \leq x \leq b\}$
$\mathbb{R}^{n\times m}$	m-columns, n-rows matrices of $\mathbb{R}$
$\mathbb{R}^n$	Vectors, identified with $\mathbb{R}^{n\times 1}$

Functions

$\circ$	Function composition
$\partial f/\partial x$	Partial derivative of f
$\dot{f}$	$df(t)/dt$

Complex numbers

$\mathrm{Re}(z)$	Real part of $z \in \mathbb{C}$
$\lvert z\rvert$	Modulus of $z \in \mathbb{C}$

Linear algebra

$x \in \mathbb{R}$	Scalar of $\mathbb{R}$
$\boldsymbol{x} \in \mathbb{R}^n$	Vector of $\mathbb{R}^n$
$\boldsymbol{M} \in \mathbb{R}^{n\times m}$	Matrix of $\mathbb{R}^{n\times m}$
x_i	i-th coordinate of $\boldsymbol{x}$
M_{ij}	(i, j)-th coefficient of $\boldsymbol{M}$
$\boldsymbol{M}^{-1}$	Inverse of $\boldsymbol{M}$
$\boldsymbol{M}^{\#}$	Moore-Penrose inverse of $\boldsymbol{M}$
$\boldsymbol{M}^\top$	Transpose of M
$\mathrm{rank}\,(\boldsymbol{M})$	Rank of $\boldsymbol{M}$
$\mathrm{range}(\boldsymbol{M})$	Image of $\boldsymbol{M}$
$\ker(\boldsymbol{M})$	Kernel of $\boldsymbol{M}$
$\det(\boldsymbol{M})$	Determinant of $\boldsymbol{M}$
$\rho(\boldsymbol{M})$	Spectrum of $\boldsymbol{M}$
$\lambda_{\max}(\boldsymbol{M})$	$\max_{\lambda\in\rho(\boldsymbol{M})} \mathrm{Re}(\lambda)$
$\lambda_{\min}(\boldsymbol{M})$	$\min_{\lambda\in\rho(\boldsymbol{M})} \mathrm{Re}(\lambda)$
$\langle \boldsymbol{a}, \boldsymbol{b}\rangle$	Scalar product $\boldsymbol{a}^\top \boldsymbol{b}$
$\langle \boldsymbol{a}, \boldsymbol{b}\rangle_{\boldsymbol{D}}$	Scalar product $\boldsymbol{a}^\top \boldsymbol{D}\boldsymbol{b}$
$\boldsymbol{I}, \boldsymbol{I}_n$	Identity matrix of $\mathbb{R}^{n\times n}$
$\mathbb{1}, \mathbb{1}_n$	All-one matrix of $\mathbb{R}^{n\times n}$
$\mathbf{0}, \mathbf{0}_n$	All-zero vector of $\mathbb{R}^n$
$\mathbf{1}, \mathbf{1}_n$	All-one vector of $\mathbb{R}^n$
$\mathrm{span}(\mathcal{S})$	Vector space spanned by $\mathcal{S}$
$\dim(\mathcal{S})$	Intrinsic dimension of $\mathcal{S}$
δ_{ij}	Kronecker delta
$\perp$	Orthogonal

Chapter 1

Introduction

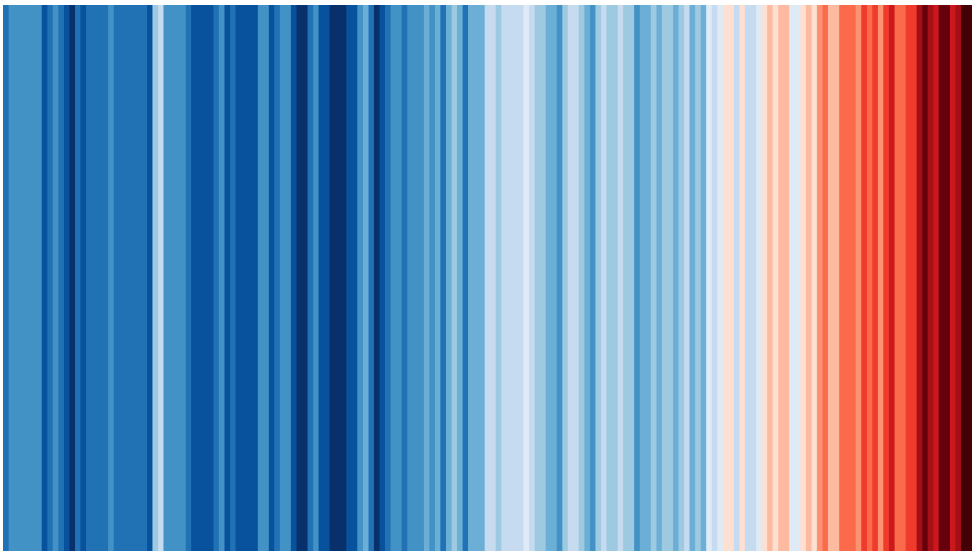


Figure 1.1: Global temperature increase from average global temperature from 1850 to 2024 [1] . The blue stripes represent years with a temperature below the average temperature between 1850 and 2024 and the red stripes represent years with a temperature above.

1.1 From Societal Imperative to Scientific Challenge

The twenty-first century is entering a decisive phase in the struggle to stabilize the Earth's climate. The global average atmospheric temperature has already risen by more than $1.3°C$ above pre-industrial levels (1750), with the past decade marking the hottest period in recorded history[2] . The question is not anymore *if* we will exceed the $2°C$ warming limit set by the Paris Agreement but *how much* we will exceed it. Extreme weather events, that were once considered rare and isolated, are now occurring with increasing frequency and intensity: heatwaves in Europe and North America [1], catastrophic floods in Pakistan [2], and record wildfires across

[1] https://showyourstripes.info/

[2] https://www.eea.europa.eu/en/analysis/indicators/global-and-european-temperatures

[1] "Heat Extremes in West Europe Increasing Faster than Simulated". Robert Vautard et al. (2023).

[2] "Causes of 2022 Pakistan Flooding: Link with China and Europe Heatwaves". Chi-Cherng Hong et al. (2023).

the Mediterranean [3], western United States (California) [4] and Australia [5], have revealed the mounting human and economic toll of a rapidly warming planet. Despite this mounting evidence and decades of coordinated international effort, from the establishment of the United Nations Framework Convention on Climate Change (UNFCCC) and the annual Conferences of the Parties (COP), to the legally binding commitments of the Paris Agreement, global greenhouse gas emissions have continued to rise, reaching new highs even in the aftermath of the COVID-19 slowdown.

This widening gap between ambition and outcome exposes a profound paradox at the heart of contemporary climate governance. Humanity possesses the scientific understanding and technological capacity to transition toward a low-carbon future: renewable energy is now cheaper than fossil fuels in much of the world, and viable pathways for decarbonization have been mapped in detail by the Intergovernmental Panel on Climate Change (IPCC) [3] [6]. Yet, collective progress remains alarmingly inadequate. This failure cannot be explained by technological or economic constraints alone. It reflects the deeper, systemic difficulty of mobilizing and sustaining coordinated action across diverse actors, institutions, and societies.

To grasp why emissions continue to rise despite widespread consensus on the need for rapid reduction, we must view decarbonization as a *socio-technical* problem, one shaped by the intertwined dynamics of policy, human behavior, and environmental feedback and not only following a leap of faith in potential technology solutions as economists often do. Technologies do not operate in isolation; they are embedded within political and cultural systems that determine how, and how quickly, they are adopted. The challenge, therefore, is not only to develop cleaner technologies, but to understand and influence the collective mechanisms that govern social cooperation, behavioral change, and institutional adaptation.

1.1.1 Why Progress Stalls: Two Key Obstacles

When we look at why progress is so slow, two interconnected obstacles, or "brakes", appear again and again.

(i) **Strategic and Economic Competition.** A stable climate is a classic global public good: while its benefits are shared by all, the costs of action fall on individual actors and are often felt immediately. This mismatch between collective benefits and individual costs creates a powerful incentive to delay action or "free-ride"[4] on the efforts of others, a dynamic often described as the **Tragedy Of the Commons (TOC)** [7]. On a national level, governments face pressure to protect local industries and jobs. At the corporate level, climate commitments are felt as a cost weighed against market competitiveness. At the individual level, the cost of climate action is also often perceived as a burden on personal convenience or financial security. This leads to a fragile situation where progress hinges on mutual expectations; each actor's willingness to commit is conditional on the perceived actions of others. This strategic interdependence makes large-scale cooperation inherently difficult to sustain, leading to the cycles of optimism and disappointment that often characterize international climate negotiations.

[3] "Climate Change Has Increased the Odds of Extreme Regional Forest Fire". John T. Abatzoglou et al. (2025).
[4] "The 2020 to 2021 California Megafires and Their Impacts on Wildlife Habitat". Jessalyn Ayars et al. (2023).
[5] "Biodiversity Impacts of the 2019–2020 Australian Megafires". Don A. Driscoll et al. (2024).
[3] https://www.ipcc.ch/report/ar6/wg1/
[6] *IPCC, 2023*. Katherine Calvin et al. (2023).
[4] Individual benefits from the actions of others without contributing to the cost of the action.
[7] "The Tragedy of the Commons". Garrett Hardin. (1968).

(ii) **Socio-Behavioral Dynamics.** The success of any climate policy ultimately depends on its social acceptance and the adaptation of collective behaviors. Public support is not based solely on scientific information; it is shaped by cultural values, group identities, and perceptions of fairness. A policy that is technically optimal may fail if it is seen as unfair or illegitimate by the public and political actors. Furthermore, human behavior is often characterized by inertia, where individuals and communities are slow to change established habits, even when more sustainable alternatives are available.

These challenges are amplified in the modern information environment. The architecture of social media can create echo chambers, reinforcing existing beliefs and deepening partisan divides. As a result, climate change is increasingly framed not as a shared scientific or policy problem, but as an issue of cultural identity or political ideology (e.g. climate denialism) [8]. This polarization wears away the common ground of shared facts needed for productive debate, making compromise more difficult. The same network effects that can rapidly spread new ideas can also fuel backlash or denial, making public opinion more volatile and susceptible to short-term events. In such a divided and fast-changing public arena, building the durable political consensus required for long-term climate strategy becomes an exceptional challenge.

These two obstacles are linked. For example, politicians are less likely to pursue ambitious climate policies if they anticipate strong public resistance or if they believe that other countries are not doing enough. At the same time, the environment itself is part of the system: events like heatwaves can change public opinion, but this feedback is often slow and affects different groups in different ways. The result is a complex system where **strategy, society, and the environment are all tied together.**

1.1.2 From Description to Explanation: The Need for a Mechanistic Understanding

While descriptions of the decarbonization challenge are common, what is needed are models that can truly *explain* its underlying mechanics. Many modeling traditions have contributed to our understanding of the decarbonization challenge. Here, we provide a concise overview, intended only to situate the reader, of three paradigms that form the conceptual toolbox of this thesis:

- **Climate-economy models** provide the essential macroeconomic and geophysical context. However, to remain tractable, they must aggregate social and political behavior, thereby abstracting away the very actor-level strategic interactions and social influence dynamics that are critical to understanding policy failures and successes [9, 10].
- **Game-theoretic models** are essential for analyzing the strategic incentives that drive cooperation or defection. They usually really simplified models of the economy and the climate and are not able to capture the complex dynamics of the real world. Additionally, they typically assume that agents' preferences are fixed, often overlooking the dynamic evolution of beliefs and social norms [11–13].

[8] "The Politicization of Climate Change and Polarization in the US". Aaron M. McCright et al. (2011).
[9] "Climate Change Policy". Robert S. Pindyck. (2013).
[10] "The Use and Misuse of Models for Climate Policy". Robert S. Pindyck. (2017).
[11] *Game Theory and International Environmental Cooperation*. Michael Finus. (2002).
[12] *Game Theory and Climate Change*. Parkash Chander. (2018).
[13] *Economics, Game Theory and International Environmental Agreements:The ca' Foscari Lectures*. Henry Tulkens. (2019).

– **Opinion dynamics models** intend to reveal how ideas and behaviors spread across social networks. Yet, they often treat the external world, such as the state of the climate or the policy environment, as a static background rather than a co-evolving part of the system.

The core limitation is that these approaches are rarely integrated analytically, leaving them unable to explain complex, non-equilibrium behaviors like abrupt opinion shifts or cycles of collective action and inaction. This recognition is not new [14], and the call to complement traditional models with paradigms like Game Theory and Agent-Based Modeling remains acutely relevant [15].

This thesis aims at fulfilling this need. The overarching scientific challenge is to build a synthesis that integrates strategic choice, the nonlinear dynamics of social influence, and feedback from a co-evolving environment. Instead of building a large, descriptive simulation, our goal has been to use the tools of game theory and dynamical systems to identify the core **mechanisms** and **tipping points** that govern the qualitative behavior of this coupled system.

1.2 A Journey into Socio-Environmental Dynamics

The analysis of complex socio-environmental systems, such as international climate policy, requires a multi-paradigm approach that builds progressively from classical strategic models to more nuanced behavioral dynamics. Our research program has been structured as a three-stage investigation, with each stage adopting a different lens to understand the mechanics of collective action and its environmental consequences.

This journey is divided into three stages which are summarized in the following:

1. **Nation-Level (Macroscopic): Strategic Analysis.** We employ the tools of **Game Theory**, to model the strategic interactions between nations at a world-wide scale in the context of international climate policy. Those strategic interactions are assumed to lead to equilibria, Nash Equilibria, and provide a framework to answer the foundational question:

   > "What are the stable policy configurations when independent nations act in their own perceived self-interest? Is it the best strategy to act in a cooperative way or to act in a selfish way? How does the climate model influence the policy decisions?"

2. **Individual-Level (Microscopic): Coupled Opinion-Environment Dynamics Analysis.** We consider the setup of **discrete-time Opinion Dynamics**, where an agent's state (e.g., its stance on climate policy) evolves based on social influence from its social network. We introduce a feedback loop where the collective opinion state drives an environmental dynamic, which in turn influences the evolution of opinions. This stage addresses the question:

   > "How do collective opinions and environmental states co-evolve over time, and what emergent phenomena, like consensus or persistent oscillations or chaos, arise from this feedback? Is polarization possible and can it be robust?"

3. **Social Network-Level (Mesoscopic): Nonlinear Dynamics Analysis.** The final step builds upon the previous trying to distinguish the different types of behaviors that can arise from the coupled system and the one emerging from the opinion dynamics model. This stage addresses the question:

   > "How does the communication and the topology of the social networks shape the potential for global consensus versus fragmentation into localized, stable

opinion clusters? From where emerge the polarization and the fragmentation?"

To maintain a concise, coherent and focused narrative, this manuscript is dedicated to a thorough exposition of the **third stage** results, which represents one of the main contributions of this thesis and mainly focus on the analysis of the nonlinear dynamics on networks. Still in aim of providing a comprehensive understanding of the mechanics of collective action and its environmental consequences, and provide a broad overview of the research realized during this thesis, we will provide a brief summary of the first two stages with associated contributions in the following sections.

1.3 First Stage: Formalizing Strategic Conflict: A Game-Theoretic Perspective

1.3.1 The Initial Question and Toolkit

The first logical step was to formalize the strategic dimension of decarbonization. The question we initially posed was:

> "Can we model the strategic decision-making of nations concerning climate policy as a game to understand the conditions for stable, self-enforcing climate agreements?"

To adress this, we turned to the established tools of **Game Theory**. This paradigm provides the canonical framework for analyzing the interactions between several decision-making agents or players using **utility functions** to capture their preferences and solution concepts like the **Nash Equilibrium** to identify stable outcomes.

1.3.2 State of the Art: A Disciplinary Fracture Between Two Worlds

The modeling literature on climate policy has largely evolved along two prolific but parallel tracks: the holistic, simulation-based approach of **Integrated Assessment Models (IAMs)** and the analytical, agent-focused approach of **Environmental Game Theory**.

Integrated Assessment Models (IAMs) On one hand, IAMs provide a "comprehensive" systems view, coupling simplified models of the economy, energy sector, and climate system to explore policy scenarios (see Figure 1.2). Their philosophical goal is to capture the essential feedbacks of the socio-environmental system. However, their complexity often renders them "black boxes," making it difficult to extract clear, analytical insights about strategic behavior. They are invaluable for scenario analysis but are not designed for the formal study of strategic equilibria.

IAMs originate from the work of Nordhaus and the **Dynamic Integrated model of Climate and the Economy (DICE)** model [16, 17], extended in the multi-agent framework in [18], which established a quantitative framework coupling the economy and the climate system. This innovation provided an essential foundation for evaluating climate policies in economic terms

[16] "To Slow or Not to Slow". William D. Nordhaus. (1991).
[17] "An Optimal Transition Path for Controlling Greenhouse Gases." William D. Nordhaus. (1992).
[18] "A Regional DGE Model of Alternative Climate-Change Strategies". William D. Nordhaus et al. (1996).

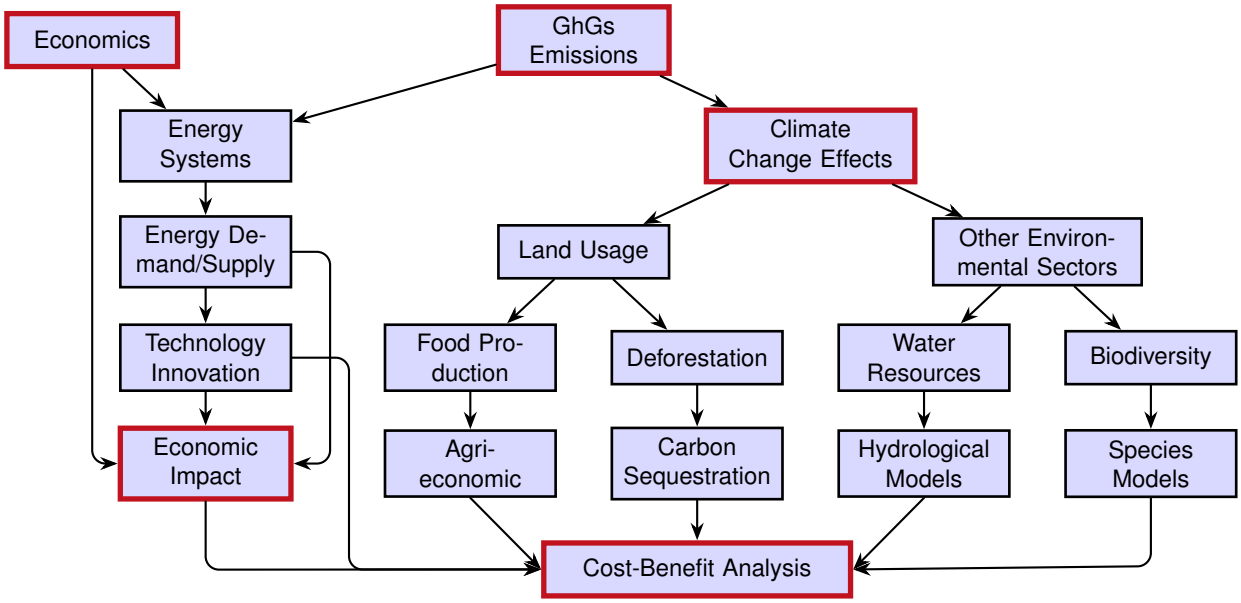


Figure 1.2: A schematic representation of an IAM possible structure illustrating their complexity and non-tractable nature.

and has since inspired numerous variants and extensions and some still being developed [19–22]. Those models are used by IPCC to generate their scenarios and studies as a part of a multi-model approach [23]. Figure 1.2 presents a possible structure of an IAM, illustrating the potential complexity and non tractable nature of such models. However, the DICE framework also exhibits well-documented limitations.

From a physical standpoint, its climate module was highly simplified and does not accurately represent the complexity of carbon-cycle dynamics or feedback processes [24–26]. The usage of the proper Simple Climate Model (SCM) is requiered to get the right impact over the environment. For instance climate model as [27–29] (only some of them are cited here for brevity) are much more accurate than the SCM used by Nordhaus initially. Those models are used by IPCC in they scenarios and studies as a part of a multi-model approach [30]. Some authors tried to incorporates better SCM into the DICE model [31] leading finally Nordhaus to use the FAIR model [32].

From an economic perspective, it relies heavily on a stylized *damage function*: a mathematical relationship linking temperature rise to economic losses. This function is both conceptually and

[19] "PAGE95". Erica L Plambeck et al. (1996).
[20] "The Marginal Damage Costs of Different Greenhouse Gases". Stephanie Waldhoff et al. (2014).
[21] *REMIND - REgional Model of INvestments and Development*. Gunnar Luderer et al. (2025).
[22] "Climate Economics Support for the UN Climate Targets". Martin C. Hänsel et al. (2020).
[23] "Long-Term Economic Growth Projections in the Shared Socioeconomic Pathways". Rob Dellink et al. (2017).
[24] "Correcting the Carbon Cycle Representation". F. Joos et al. (1999).
[25] "How Well Do IAMs Represent Non-CO2 Radiative Forcing?" Mathijs J. H. M. Harmsen et al. (2015).
[26] "Are Economists Getting Climate Dynamics Right and Does It Matter". Simon Dietz et al. (2021).
[27] "Carbon Dioxide and Climate Impulse Response Functions". F. Joos et al. (2013).
[28] "Transient Climate Response in a Two-Layer Energy-Balance Model. Part I". O. Geoffroy et al. (2013).
[29] "FaIRv2.0.0". Nicholas Leach et al. (2021).
[30] "Overview of CMIP6". V. Eyring et al. (2016).
[31] "Towards a FAIR-DICE IAM". Timm Faulwasser et al. (2018).
[32] "Policies, Projections, and the Social Cost of Carbon". Lint Barrage et al. (2024).

empirically fragile [33, 34]. Because the true shape of this function cannot be known before the impacts of climate change fully materialize [9, 10], such models risk giving a false sense of precision in policy evaluation. The only assumption that seems to be common in the literature is the convexity of the damage function with respect to the temperature but the function in itself is still a subject of debate [35, 36]. Moreover, an additional limitation is the presence of a discounting factor is a major limitation of the model [37] as a small change in the discounting factor can lead to completely different policy recommendations.

Those two elements, the climate model and the damage function, are central to get the right quantification of impact on the economy. For instance, in the initial stage of the model, and with the SCM used by Nordhaus, yields an optimal trajectory for the emissions corresponding to an increase of $4.5°$C in the global average temperature which is far from the current understanding of the climate system.

Environmental Game Theory On the other hand, Environmental Game Theory provides the analytical rigor to study strategic interactions. The paradigmatic model in this literature, following the work of Finus, Chander, Tulkens, and others [11–13], represents an agent's utility u_n as a trade-off between the benefits (B_n) of its own emissions a_n and the damage function (D_n) resulting from the total sum of all emissions $\sum_{k=1}^{N} a_k$:

$$u_n(a_1, \ldots, a_N) = \underbrace{B_n(a_n)}_{\text{Benefit}} - \underbrace{D_n\Big(\sum_{i=1}^{N} a_i\Big)}_{\text{Damage}}. \qquad (\text{CO}_2 \text{ Game})$$

This raises a key question:

> "Why did this elegant but physically abstract model become the canonical framework, especially when the IAM tradition had already established the importance of dynamic climate feedback?"

The answer appears to lie in disciplinary priorities and a historical trade-off. As this branch of game theory flourished, its primary goal was to achieve clean, provable results about strategic incentives [11–13] and coalition stability [38, 39]. The formulation in (CO_2 Game) is mathematically convenient, allowing for a deep analysis of equilibria.

However, this analytical clarity was purchased at a high price. The abstraction from geophysical complexity led to two major issues. First, it created a wide gap between theoretical elegance and empirical validation, as the simplified damage functions were nearly impossible to calibrate against real-world climate processes [9]. Second, it led to theoretically pessimistic and sometimes paradoxical conclusions, such as Barrett's "coalition paradox", which famously suggested that the conditions for forming large, stable climate coalitions were almost impossible to meet [40].

[33] "Additive Damages, Fat-Tailed Climate Dynamics, and Uncertain Discounting". Martin Weitzman. (2009).
[34] "What Is the "Damages Function" for Global Warming?" Martin L. Weitzman. (2010).
[35] "The Economic Impacts of Climate Change". Richard S. J. Tol. (2018).
[36] "Social Cost of Carbon Estimates Have Increased over Time". Richard S. J. Tol. (2023).
[37] "The Politics of Climate Denialism and the Secondary Denialism of Economics". Clifford William Cobb. (2025).
[38] "Incentives and Stability of International Climate Coalitions". Valentina Bosetti et al. (2013).
[39] "The Efficacy of International Environmental Agreements". Michael Finus et al. (2021).
[40] "Self-Enforcing International Environmental Agreements". Scott Barrett. (1994).

In a sense, a disciplinary path dependency was established: while IAMs pursued ever-greater descriptive complexity, much of the game-theoretic community prioritized analytical purity. The valuable lessons from climate science about nonlinear dynamics and feedback were not fully re-integrated into the strategic models. To mention in [13], simulation are conducted but with DICE model which was already outdated both in terms of climate modeling and in terms of damage function.

Our first contribution, detailed in the next section, can be seen as a direct response to this divergence, a first step toward healing this disciplinary fracture by re-embedding a dynamic climate model within a tractable game-theoretic framework.

1.3.3 Our Contributions: A Construction of a Socio-Environmental Game

The body of work detailed in this section represents a multi-stage research program developed in close and synergistic collaboration with my colleague, Dr. Bouchra Mroué. Our joint inquiry was designed to systematically construct a novel game-theoretic framework for analyzing climate policy, with each publication marking a deliberate step in this intellectual progression. The following subsections summarize the core contribution of each key paper in chronological order to delineate this development. While this overview provides the conceptual architecture of our approach, the reader is directed to the cited publications or her thesis [41] for the complete mathematical formalism, proofs, and detailed numerical analyses that underpin our findings.

The Foundational Step: Re-Embedding Climate Physics into Game Theory

[42] Mroué, B., Couthures, A., Lasaulce, S., & Morărescu, I.-C. (2023). Une approche par la théorie des jeux du problème du choix des niveaux nationaux d'émissions de CO2. XXIXème Colloque Francophone de Traitement Du Signal et Des Images, GRETSI 2023. https://gretsi.fr/data/colloque/pdf/2023_mroue1148.pdf

[43] Couthures, A. (2025). AnthonyCouthures/CO2-Game [Jupyter Notebook]. https://github.com/AnthonyCouthures/CO2-Game (Original work published 2022)
Documentation: https://anthonycouthures.github.io/CO2-Game/

The first step was to reformulate the core strategic interaction under an accurate physical model. We departed from the physically abstract damage function of classical environmental games, $D_n(\sum a_k)$, which treats emissions as a direct proxy for harm. Instead, we posited that a nation's utility is contingent on the **global atmospheric temperature (θ^{AT})**, a nonlinear and dynamic output of a coupled geophysical model from a weighted common damage function: $w_n D(\theta^{\mathrm{AT}})$ for some weights $w_n > 0$.

The central theoretical contribution of this initial work was to prove that, despite the introduction of this complex, state-dependent feedback loop, the resulting formulation constitutes a **weighted potential game**. This result, established within the framework of Monderer and Shapley [44], is the building block of our research program on climate change and game theory. It guarantees the existence of a pure-strategy Nash equilibrium, ensuring that the model is not an intractable "black box" but an analytically sound framework amenable to rigorous equilibrium analysis and simulation guarantees. Moreover, the potential game structure allows an easy and fast numerical search for Nash equilibria using the Best Response Dynamics.

[41] Bouchra Mroué. *"Decarbonization Strategies: A Game Theory Perspective"*. PhD thesis. Université de Lorraine, Apr. 2024.
[44] "Potential Games". Dov Monderer et al. (1996).

From Existence to Expression: Deriving Tractable Solutions

[45] Mroué, B., Couthures, A., Lasaulce, S., & Morărescu, I.-C. (2024). A Quadratic Static Game Model for Assessing the Impact of Climate Change. In H. Cherifi, L. M. Rocha, C. Cherifi, & M. Donduran (Eds.), Complex Networks & Their Applications XII (pp. 383-394). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-53499-7_31

The next logical task was to characterize the Nash equilibrium. In our subsequent paper, we specialized the model to the analytically convenient case of quadratic benefit and damage functions as suggested in [34]. This allowed us to move from an abstract existence proof to a concrete analytical solution, the unique pure Nash Equilibrium.

We derived a **closed-form expression for the unique pure Nash Equilibrium** and analytical conditions for the existence and uniqueness of the Nash equilibrium in the game. This was a needed step forward, as it rendered the model's predictions explicit and interpretable. It enabled us to conduct the first quantitative analyses demonstrating precisely how the perceived severity of the damage function, a key policy variable, directly and predictably alters the strategic emission choices of nations.

Incorporating Socio-Political Realism: The Networked Imitation Game

[46] Mroué, B., Couthures, A., Lasaulce, S., & Morărescu, I.-C. (2025). On the impact of cross-country imitation on climate change: A game-theoretical analysis. IFAC Journal of Systems and Control, 32, 100309. https://doi.org/10.1016/j.ifacsc.2025.100309

Climate policy is not formed in a political vacuum; it is shaped by international pressures, alliances, and strategic mimicry. To capture this reality, this work introduced a richer layer of socio-political structure. We augmented the utility function with an **imitation term**, effectively modeling the strategic interactions as a game played on a network. This work is the first to study the impact of network topology on the equilibrium of the game.

This formulation allowed us to shortly investigate how **network topology** influences collective outcomes. We explicitly compared scenarios of a fully-connected, globalized world with one of fragmented, competing blocs. The key contribution here was to provide a formal, quantifiable basis for the intuition that greater interdependence enforces better climate outcomes. We demonstrated that a more connected influence graph systematically leads to lower equilibrium temperatures, thereby highlighting the critical role of inclusive international diplomacy in achieving decarbonization goals.

Bridging Static Analysis and Dynamic Policy: Planning and Adaptation

- Couthures, A., Mroué, B., Lasaulce, S., & Morărescu, I.-C. A static game model for climate change analysis and strategy design. *(Preprint)* In appendix

The final and most comprehensive stage of our work addresses the inherently long-term nature of climate policy, bridging the gap between static equilibrium analysis and dynamic strategy design. This work recognizes that real-world policy is not a myopic, one-shot game but an ongoing process of long-term planning, commitment, and adaptation.

Theoretically, our first contribution is to demonstrate that the potential game structure of the static model is robustly preserved when extended to a multi-stage *planning game*. This is a non-trivial extension that underpins the analytical tractability of the long-horizon problem, guaranteeing the existence of pure Nash equilibria. Furthermore, we provide deep insights into the structure of the equilibrium set itself. Specifically, we prove an ordering property for the Nash equilibria, a result related to the concept of supermodularity in games. This establishes that if one equilibrium involves a player emitting less than in another equilibrium, then all other players must also emit less or the same, preventing unordered or paradoxical outcomes.

From a policy perspective, a central contribution is our detailed numerical investigation into the conditions for cooperation. We demonstrate the critical role and frequent inadequacy of conventional damage functions. Our results show that simple quadratic damages, while convex and common in the literature, are often **far too optimistic to induce the level of cooperation required to meet ambitious climate targets**. Achieving significant decarbonization within the model requires assuming more severe, non-linear damage functions, highlighting this function's role as a critical lever for policy analysis. Finally, our comparative analysis shows that the *receding horizon strategy*, which allows for adaptive re-planning, systematically yields superior climate outcomes, namely lower equilibrium temperatures, compared to a static, open-loop planning commitment.

To conduct this analysis, we formalized and systematically compared three distinct temporal decision-making frameworks:

- The *repeated one-shot game*, representing myopic, short-term decision-making.
- The *planning game*, where agents commit to an optimal emissions trajectory over a long horizon (e.g., to 2100).
- The *receding horizon strategy*, a more adaptive approach from control theory where agents iteratively re-optimize their plans as time progresses.

The contribution of this paper is a detailed numerical exploration of the trade-offs between these strategies, benchmarked against the cooperative Social Optimum. It quantifies the important **value of foresight**, demonstrating that longer planning horizons systematically yield superior climate outcomes. By analyzing these different solution concepts, this work provides a framework for understanding the strategic implications of different international policy architectures, from short-term pledges to long-term, binding treaties. Other insights are, for example, the non stability of the Grand coalition as well as the non uniqueness of the emission trajectories for identical economic outcomes.

1.3.4 Insight and Pivot to the Next Stage

This game-theoretic investigation provided a robust, analytically and numerically grounded framework for understanding stable strategic outcomes in the face of environmental feedback and network influence. We successfully identified the conditions for equilibrium and quantified how policy levers, such as damage perception and planning horizons, shape these stable points.

However, a profound limitation became apparent. The first stage has focused on the steady state behavior and consequently, our analysis could not capture the *temporal dynamics of collective behavior itself*. It identifies strategically stable attracting landing points, the Nash equilibria, but is uniformative on the transient journey. It cannot, by design, explain the emergence of complex, non-equilibrium phenomena such as the persistent oscillations observed in public opinion, the sudden, system-wide shifts in policy consensus, or the stable coexistence of polarized clusters. These emergent behaviors are the signature of complex feedback systems operating out of equilibrium.

A pivotal change in our research perspective has been triggered by the realization that a static equilibrium view, although sophisticated, was by design insufficient to capture the richness of socio-environmental dynamics. We moved from the analysis of **strategic choice** to the analysis of the **dynamic evolution of collective states**. Thus, we moved from the analysis of **strategic choice** to the analysis of the **dynamic evolution of collective states**. This transition required a new set of tools and a new conceptual lens, launching the second stage of this research.

1.4 Second Stage: Interplay of Environment and Opinion Dynamics

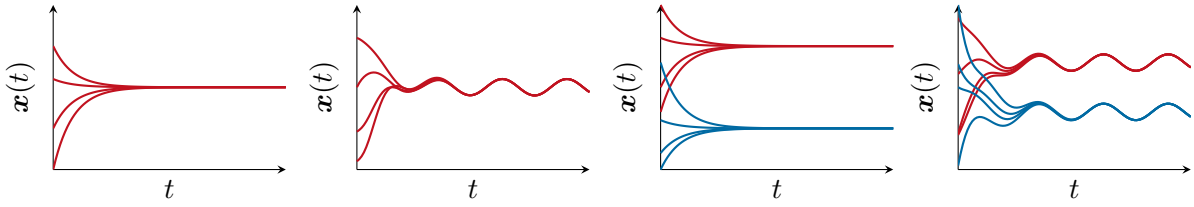
1.4.1 The Evolving Question and Toolkit

The limitations of Nash equilibrium analysis made it clear that a deeper understanding requires capturing the temporal evolution of opinions and environmental states. The question thus became:

> "What happens to collective behavior when we explicitly couple the social dynamics of opinion formation with a co-evolving environmental state?"

To answer this, we synthesized tools from two key domains: **Nonlinear Dynamical Systems**, which brings the concepts of dynamical stability, attractors, and bifurcations; and the literature on **Opinion Dynamics**, which offers formalisms for modeling how beliefs evolve over social networks.

Figure 1.3 presents the observed dynamic behaviors of the coupled system obtained from the start of this second stage of research. The four panels represent the four possible dynamic behaviors of the system: constant consensus, oscillating consensus, constant polarized, and oscillating polarized.



(a) Constant consensus (b) Oscillating consensus (c) Constant polarized (d) Oscillating polarized

Figure 1.3: Initial observed dynamic behaviors of the coupled system.

1.4.2 State of the Art: The Rise of Coupled Socio-Environmental Models

While the general idea that social and environmental systems co-evolve is not new, its formal, analytical modeling has undergone a dramatic acceleration in just the past few years. At the outset of this doctoral research, the landscape was considerably less codified than it is today. Much of the key literature that now defines this emerging field was published in parallel with the work presented in this thesis. This section looks at these recent developments as a map of active investigations and not as a fixed backdrop. It illustrates how scholars from several related disciplines are beginning to probe the feedback loop between human behavior and environmental dynamics.

Feedback-Evolving Games and Eco-Evolutionary Dynamics A foundational stream of work emerged from mathematical biology and **evolutionary game theory**. The seminal contribution of Weitz et al. [47] introduced the concept of a "feedback-evolving game," demonstrating that when a population's strategic choices (e.g., cooperation vs. defection) both depend on and influence an environmental state (e.g., a shared resource), the system can spontaneously produce sustained oscillations: an "oscillating tragedy of the commons". This foundational idea

[47] "An Oscillating TOC in Replicator Dynamics with Game-Environment Feedback". Joshua S. Weitz et al. (2016).

was generalized by Tilman et al. [48] into a broader eco-evolutionary framework, which clarified the conditions under which such systems converge to stable equilibria versus exhibiting cycles, highlighting the role of timescale separation between the social and environmental dynamics. Complementing these evolutionary approaches, recent studies have extended the framework to include **Mean-Field Games** with switched systems [49] and accounted for irrational behaviors via **Prospect Theory** [50]. This analytical approach has been further refined with tools from control theory to rigorously prove the existence and stability of limit cycles via bifurcation analysis [51], and has matured to the point of a systematic classification of all possible dynamic outcomes in simple two-strategy games [52].

Incorporating Structural Complexity: Space and Networks The initial models typically assumed "well-mixed" populations, where any individual interacts with any other. Subsequent research has added important layers of structural realism. Lin & Weitz extended the feedback-evolving game to a spatial context, showing the emergence of complex spatiotemporal patterns and anti-phase oscillations, where different regions of the system cycle out of sync [53]. More recently, Betz et al. introduced an explicit network structure with community divisions, demonstrating that the feedback loop can induce multistability and anti-synchrony between the connected communities [54]. This line of work proves that network topology is a key determinant of the system's macroscopic behavior.

The Epidemiological Approach: Modeling Social Contagion A second influential paradigm borrows its tools from **mathematical epidemiology**, treating social phenomena like rumor propagation or behavioral adoption as a contagious process. This approach uses compartmental models (e.g., SIR: Susceptible-Infected-Recovered) to describe the flow of a population between a small number of discrete states. This framework is particularly well-suited to modeling misinformation, as demonstrated by Satheesh Kumar et al., who developed a coupled socio-climate model where a climate-denying rumor spreads through a population divided into susceptibles, hesitators, believers, and rejectors [55]. The strength of this approach lies in its ability to leverage a well-understood and analytically rich toolkit for studying propagation thresholds and dynamics. The mathematical depth of this field is further illustrated by work on reaction-diffusion models for environmentally driven infections, which incorporate spatial heterogeneity and provide rigorous conditions for the persistence or extinction of the "infection" [56]. The primary limitation of this paradigm, however, is its reliance on a few discrete behavioral categories (e.g., believer/rejector), which may not capture the full, continuous spectrum of human opinion on complex issues like climate change. It is important to note that this modeling paradigm is not limited to climate science; it has been applied with great success in fields like epidemiology to model the co-evolution of disease prevalence and vaccination behavior [57].

[48] "Evolutionary Games with Environmental Feedbacks". Andrew R. Tilman et al. (2020).
[49] "MFG for Collective Decision-Making in Honeybees via Switched Systems". Leonardo Stella et al. (2022).
[50] "The Impact of Irrational Behaviors in the OPD with Game-Environment". Leonardo Stella et al. (2023).
[51] "Limit Cycles Analysis and Control of Evolutionary Game with Environment". Lulu Gong et al. (2022).
[52] "A Complete Classification of Evolutionary Games with Environmental Feedback". Hiromu Ito et al. (2024).
[53] "Spatial Interactions and Oscillatory Tragedies of the Commons". Yu-Hui Lin et al. (2019).
[54] "Evolutionary Game with Environmental Feedback with Two Communities". Katherine Betz et al. (2024).
[55] "Climate-Denying Rumor Propagation in a Coupled Socio-Climate Model". Athira Satheesh Kumar et al. (2025).
[56] "Dynamics in a Reaction-Diffusion Epidemic Model with Environment". Ning Wang et al. (2022).
[57] "Behavioral Vaccination Policies and Game-Environment in Epidemic Dynamics". Ariful Kabir et al. (2023).

Agent-Based Models and Coupled Opinion Dynamics A third and parallel research stream approaches the problem from the perspective of social physics and **Agent-Based Modeling (ABM)**. Here, the focus is less on stylized analytical models and more on building bottom-up simulations from the explicit, micro-level rules governing agent behavior. This approach allows for immense heterogeneity and has been used to explore a wide array of feedback mechanisms. For instance, models where agents' opinions co-evolve with their spatial positions have been used to demonstrate how feedback can drive clustering and consensus [58]. In the specific context of climate change, a growing number of sophisticated models now couple social processes (such as the perception of climate risks [59], social learning, and political polarization) with simplified climate models to generate realistic emission pathways [60]. This research is invaluable for its descriptive richness and has been applied to study pressing issues like the impact of climate denialism on environmental degradation [61, 62]. The versatility of this paradigm even extends to the interface with game theory and network science; for example, Bernardo et al. model climate negotiations as a *social power game* [63], while Zino, Cao, and collaborators have developed multilayer frameworks to rigorously analyze the coevolution of opinions and collective decision-making [64], applying mean-field approximations to capture the interplay between climate impact and population behavior [62, 65]. The strength of ABMs is their ability to capture real-world complexity; their primary trade-off is often a loss of analytical tractability. While these models can show *that* a phenomenon occurs, their complexity can make it difficult to formally prove *why* it occurs or to derive the precise mathematical mechanisms that govern it.

The Identified Gap This survey of a rapidly evolving field reveals a clear analytical gap. The most rigorous analytical work has been concentrated in paradigms (evolutionary game theory, epidemiology) whose behavioral assumptions can be limiting for studying human opinion. Conversely, the models that employ more nuanced, continuous opinion dynamics are often large-scale simulations where formal analytical characterization is sacrificed for descriptive realism. When our research began, there was a pressing need for a framework that could synthesize the strengths of these approaches: a model grounded in a **continuous representation of the opinion state**, yet structured enough to permit the rigorous **analytical investigation of its dynamic mechanisms** via bifurcation theory but not too complex to be intractable. It is precisely this gap that our work was designed to fill without even knowing it before starting this research.

1.4.3 Our Contributions: Uncovering the Mechanisms of Dynamic Behavior

We sought to build and analyze a model that was simple enough to be analytically tractable yet rich enough to capture the essential nonlinearities of both social interaction and environmental feedback. Our first contribution was to develop a hybrid coupled model that combines the strengths of the analytical and simulation approaches.

[58] "Feedback Loops in OD of ABMs with Multiplicative Noise". Nataša Djurdjevac Conrad et al. (2022).
[59] "On Modeling Collective Risk Perception via Opinion Dynamics". Lorenzo Zino et al. (2024).
[60] "Determinants of Emissions Pathways in the Coupled Climate–Social System". Frances C. Moore et al. (2022).
[61] "Overcoming Inaction". Tabea Hoffmann et al. (2024).
[62] "A Behaviour–Environment Model for the Impact of Climate Change Denial". Kathinka Frieswijk et al. (2025).
[63] "Quantifying Leadership in Climate Negotiations". Carmela Bernardo et al. (2023).
[64] "A Two-Layer Model for Coevolving Opinion Dynamics and Collective Decision". Lorenzo Zino et al. (2020).
[65] "Modeling the Co-evolution of Climate Impact and Population Behavior". K. Frieswijk et al. (2023).

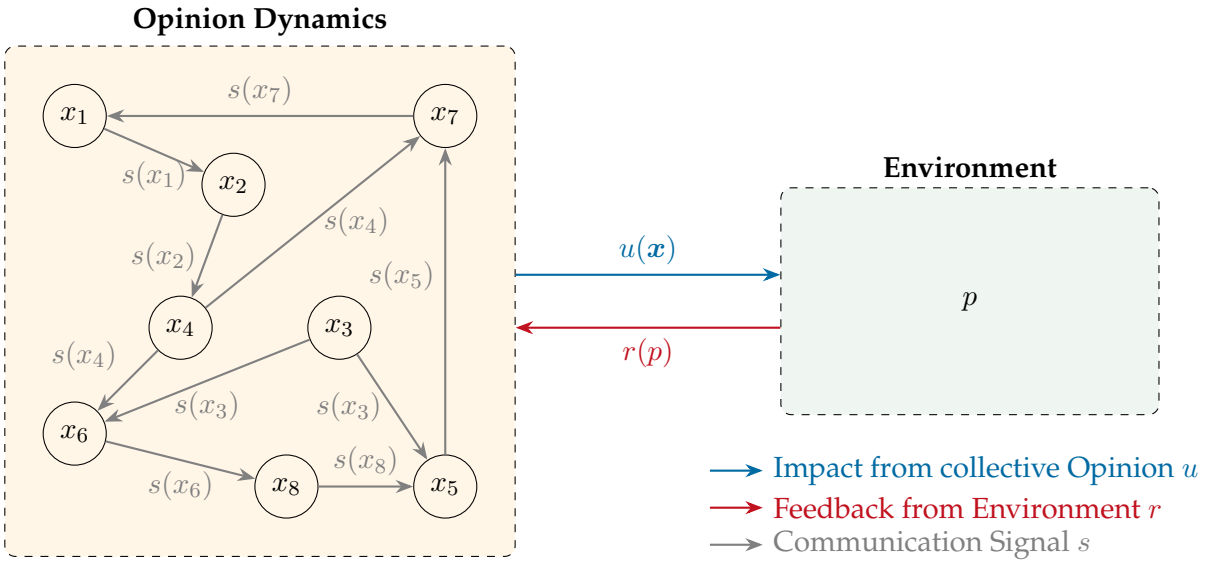


Figure 1.4: Diagram of the coupled dynamics. The opinion state of each agent is represented by the scalar $x_i(t)$ and the environmental state is represented by the scalar $p(t)$. The opinion state of each agent is influenced by the environmental state through the function $r(p(t))$ and the environmental state is influenced by the opinion state of each agent through the function $u(\boldsymbol{x}(t))$.

Initial Exploration: A discrete-time coupled model

> [66] Couthures, A., Mongaillard, T., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (2024). Analysis of a Continuous Opinion and Discrete Action Model Coupled with an External Dynamics. 2024 European Control Conference (ECC), 3392–3397. https://doi.org/10.23919/ECC64448.2024.10591291

Our first step was to construct and analyze a hybrid system that coupled a continuous-opinion/discrete-action (CODA) model [67] with a co-evolving environmental state, as shown in Figure 1.4. This discrete-time model, introduced at ECC 2024, was instrumental in revealing the system's rich dynamic potential. We demonstrated that the system's behavior depends critically on the coupling strength: the "trust" agents place in the environmental signal β. The discrete-time model can be expressed in a stylized form as:

$$\text{(Opinion)}(k+1) = \text{(Opinion)}(k) + \beta \times \text{(Env. feedback)}(k) + (1-\beta) \times \text{(Soc. influence)}(k)$$

When this coupling was weak $\beta \simeq 0$, the system's behavior recovered the classical outcomes of the uncoupled CODA model: the network self-organized into stable, **robust polarized clusters** depending on initial conditions. However, the central discovery was that when the coupling strength surpassed a critical threshold, $\beta \geq 0.5$, these stable equilibrium clusters were destroyed. The system then entered a new dynamic regime characterized by persistent oscillations, including **limit cycles** and even **chaotic dynamics**. This work provided the definitive "proof of concept" that the feedback loop itself was a mechanism capable of altering the system's asymptotic behavior, moving it beyond the static clustering paradigm.

[67] "Continuous Opinions and Discrete Actions in Social Networks". N. R. Chowdhury et al. (2016).

A Comprehensive Analysis of the Discrete-Time Coupled System

[68] Couthures, A., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (2024). Analysis of an opinion dynamics model coupled with an external environmental dynamics. Chaos, Solitons & Fractals, 189, 115719. https://doi.org/10.1016/j.chaos.2024.115719

Building upon these initial findings, we conducted a deeper, more comprehensive investigation of this coupled discrete-time system, leading to a journal publication. This work formalized the prey-predator analogy for the observed oscillations for a wide range of communication signals and provided a more systematic characterization of the different dynamic regimes. A key analytical step in this work was to first establish the **forward invariance of the synchronization manifold** $\mathcal{S}$: the subspace where all agents share an identical opinion. By proving that trajectories starting on this manifold remain on it, we could rigorously reduce the N-dimensional problem to a more tractable two-dimensional system governing the evolution of a single collective opinion and the environmental state. This analysis solidified the numerical observations from our initial exploration, formally delineating the parameter spaces that lead to either weak coupling (where stable, polarized clusters persist) or strong coupling (where convergence to a steady state is hampered, leading to limit cycles or chaos).

Bifurcation Theory in a Continuous-Time Model

[69] Couthures, A., Bizyaeva, A., Satheeskumar Varma, V., Franci, A., & Morărescu, I.-C. (2025). Bifurcation analysis of an opinion dynamics model coupled with an environmental dynamics. (preprint accepted to CDC2025) arXiv:2504.03419. https://doi.org/10.48550/arXiv.2504.03419

The discovery and characterization of these phenomena in discrete time made it imperative to seek a more rooted, mechanistic explanation. This led to the central analytical work of this thesis stage: the development and analysis of a fully continuous-time model. This shift from discrete iterations to a continuous flow allowed us to leverage the rigorous tools of **bifurcation theory**. Building on the insight of the forward-invariant synchronized manifold, we were able to formally treat the trust parameter as a bifurcation parameter and analytically prove the existence of the two critical transitions that underpin the observed dynamics:

- A **Pitchfork Bifurcation**, which marks the point where a single, unified consensus state loses stability and splits into two distinct, stable, and polarized opinion states. This provides a formal mechanism for the emergence of societal bistability.
- A **Hopf Bifurcation**, which marks the point where a stable equilibrium loses its stability and gives rise to a stable, oscillating limit cycle. This provides a formal mechanism for the emergence of sustained, periodic cycles of public attention and collective behavior.

This final step moves beyond observing the phenomena to providing a rigorous, mathematical explanation of their origin, confirming that these complex behaviors are properties of the coupled system's structure.

1.4.4 Insight and Pivot to the Next Stage

This stage of research was a critical turning point. It moved the analysis beyond static equilibria and demonstrated that the feedback loop between a social system and its environment is a strong driver for generating complex, dynamic patterns. The formal characterization of these bifurcations provided not just a description of what happens, but a rigorous explanation for *why* and *when* it happens.

However, the analysis centered on the "synchronized manifold" (the case where all agents share a common opinion). This simplification was necessary to make the bifurcation analysis

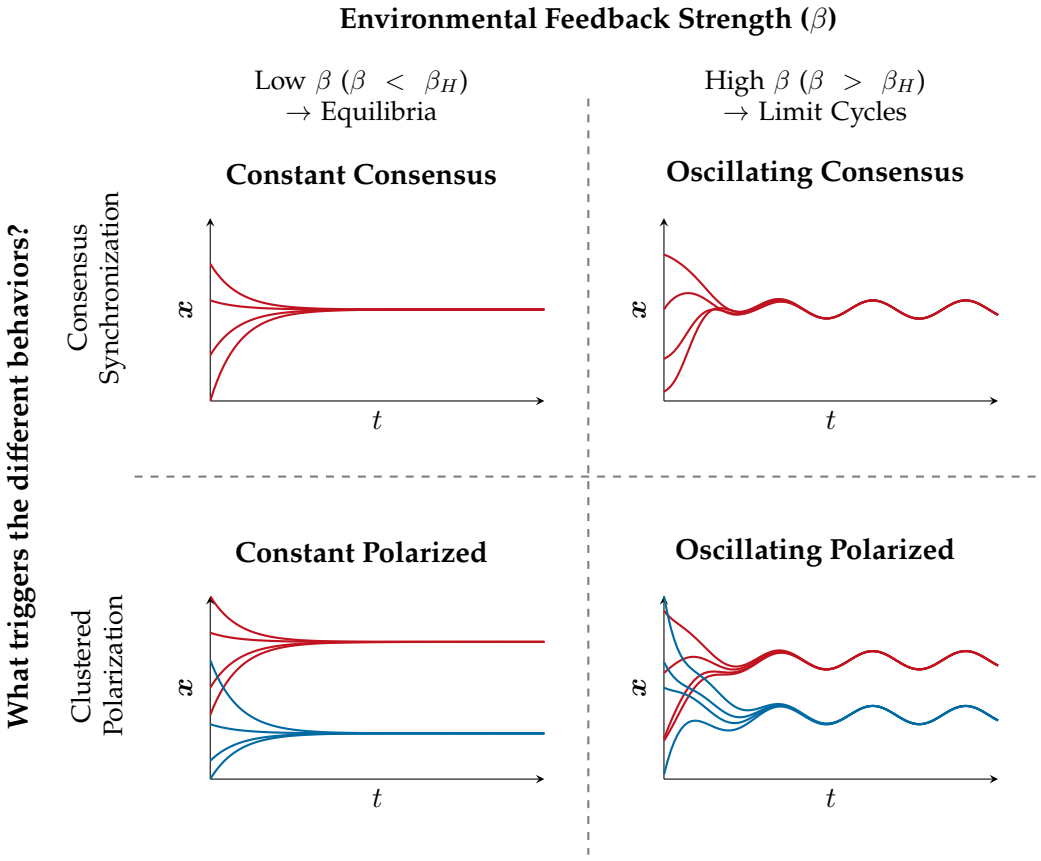


Figure 1.5: A Conceptual Phase Diagram of Socio-Environmental Dynamics with a missing element.

tractable, but it left a question unanswered: what happens in a large, heterogeneous network where agents do not synchronize? How do clusters form, maintain their coherence, and interact in the presence of nonlinear signals? To answer this, we needed a framework that could handle disagreement and partial synchronization with the same level of rigor. This required a deep dive into the foundational properties of the opinion dynamics model itself, which launched the final stage of this research.

1.5 Third Stage: The Fundamental Mechanisms of Polarization

The discovery of these complex dynamics made it imperative to move beyond numerical observation to rigorous analytical characterization. We understood how oscillations and chaos emerged from the coupled system as the horizontal axis on the phase diagram of Figure 1.5, but we where still lacking a formal understanding of the mechanisms that govern the transitions between consensus and polarization.

> "How does the network structure and the communication mechanisms affect the transitions between consensus and polarization?"

Answering this question would allow us to complete the phase diagram of Figure 1.5 and to understand the precise mathematical mechanisms that govern the transitions between consensus

and polarization.

1.5.1 State of the Art: Models of Opinion Dynamics with community structure

A central theme in the study of multi-agent systems is the emergence of collective behavior from local interactions. The foundational model in this domain, particularly for opinion dynamics, is the linear averaging protocol introduced in the seminal work of [70] extended in [71]. In its discrete-time form, the opinion $x_i(k)$ of an agent i evolves according to a weighted average of its neighbors' opinions. This can be expressed compactly for the entire network as:

$$\boldsymbol{x}(k+1) = \boldsymbol{W}\boldsymbol{x}(k), \qquad \text{(DeGroot Model)}$$

where $\boldsymbol{x}(k) \in \mathbb{R}^N$ is the vector of all agent states and $\boldsymbol{W}$ is a row-stochastic matrix (i.e. the sum of each row is 1) of interpersonal influences. The continuous-time analogue, which became a cornerstone of the control theory approach to networked systems, is described by the Laplacian dynamics [72]:

$$\dot{\boldsymbol{x}}(t) = -\boldsymbol{L}\boldsymbol{x}(t), \qquad \text{(Laplacian Model)}$$

where $\boldsymbol{L}$ is the graph Laplacian.

For decades, research has rigorously characterized the conditions under which these dynamics guarantee convergence to a global consensus. It is now well-established that for a network graph that is strongly connected (in the directed case) or connected (in the undirected case), all agents will asymptotically agree on a single value [73, 74]. The rate of this convergence is tied to the spectral properties of the system matrix, specifically, the second-largest eigenvalue modulus of W or the algebraic connectivity (the smallest non-zero eigenvalue) of L [72, 75]. This elegant framework, built upon the bedrock of matrix theory and algebraic graph theory, successfully explains how local averaging can produce global coordination even under switching topologies and delays [76–78].

However, the prediction of global agreement, while mathematically robust, is often at odds with observed social phenomena such as political polarization, cultural fragmentation, and the persistence of minority opinions [79]. The classical models (DeGroot Model) and (Laplacian Model), by their very linear construction, cannot account for the formation of stable opinion clusters or persistent disagreement. This discrepancy has motivated a series of critical extensions to the foundational model, each relaxing a key idealization to introduce richer, non-consensus dynamics [80]. The realy detailed tutorial on opinion dynamics [81] and [82] also provides a good overview of the extensions to the foundational model discussed in the following paragraphs and provides a lot of references.

[70] "A Formal Theory of Social Power." John R. P. French. (1956).
[71] "Reaching a Consensus". Morris H. Degroot. (1974).
[72] "Consensus and Cooperation in Networked Multi-Agent Systems". Reza Olfati-Saber et al. (2007).
[73] "A Necessary and Sufficient Condition for Reaching a Consensus". Roger L. Berger. (1981).
[74] *Graph Theoretic Methods in Multiagent Networks*. Mehran Mesbahi et al. (2010).
[75] "Fast Linear Iterations for Distributed Averaging". Lin Xiao et al. (2004).
[76] "Coordination of Groups of Mobile Autonomous Agents". A. Jadbabaie et al. (2003).
[77] "Consensus Seeking under Dynamically Changing Interaction Topologies". Wei Ren et al. (2005).
[78] "Convergence in Multiagent Coordination, Consensus, and Flocking". V.D. Blondel et al. (2005).
[79] "Statistical Physics of Social Dynamics". Claudio Castellano et al. (2009).
[80] "Models of Social Influence". Andreas Flache et al. (2017).
[81] "A Tutorial on Modeling and Analysis of Dynamic Social Networks. Part I". Anton Proskurnikov et al. (2017).
[82] "A Tutorial on Modeling and Analysis of Dynamic Social Networks. Part II". Anton Proskurnikov et al. (2018).

Stubbornness or Prejudice One of the most direct approach to model opinion clustering is the effect of **stubbornness or prejudice**. The Friedkin-Johnsen model, for instance, modifies the (DeGroot Model) dynamics by assuming that agents retain a degree of commitment to their initial opinions [83]:

$$\dot{\boldsymbol{x}}(t) = (1-\alpha)\boldsymbol{W}\boldsymbol{x}(t) - \boldsymbol{x}(t) + \alpha\boldsymbol{x}(0), \qquad \text{(FJ Model)}$$

where $\alpha \in [0,1]$ is a parameter that controls the degree of anchoring to the initial opinions. This anchoring prevents opinions from being fully dragged by social influence, resulting in a steady-state distribution of opinions that reflects both the network topology and the agents' initial beliefs, thus preserving a diversity of opinions rather than collapsing to a single point [84, 85].

Bounded Confidence Another direct approach to modeling opinion clustering is the introduction of **bounded confidence**. In these nonlinear models, pioneered by [86] and [87], an agent i only considers the opinions of neighbors j that are sufficiently close to its own, i.e., if $\|x_i - x_j\| \leq \epsilon$ for some confidence threshold ϵ. This homophily-driven interaction rule naturally partitions the network: agents cease to interact across large opinion gaps, leading to the emergence of multiple, distinct, and internally cohesive opinion clusters. Constraining the convergence speed of opinion between agents allows [88] to characterize strong communities satisfying a spectral graph theoretic condition.

Antagonistic Interactions Another important extension incorporates **antagonistic interactions**. The classical models implicitly assume all interactions are cooperative (positive weights). By contrast, the framework of signed graphs allows for both positive (trust, friendship) and negative (distrust, animosity) relationships. The seminal work of [89, 90] demonstrated that the system's fate depends on the network's "structural balance," a concept rooted in social psychology. If the graph is structurally balanced, the network does not reach a global consensus but rather a "bipartite consensus," where agents converge to two opposite opinion values, forming two polarized factions. If the graph is unbalanced, opinions are typically driven to zero, representing a failure to sustain any coherent belief. [91]

Perception-Based and Action-Driven Influence A final, highly active branch of research introduces non-consensus dynamics by breaking the foundational assumption that agents have direct, linear access to their neighbors' internal opinions. Instead, influence is mediated through a **perception layer**, representing the cognitive and communicative processes that transform a latent opinion into an influential signal. This has led to two distinct but converging lines of inquiry: one focusing on nonlinear opinion reinforcement and another on the consequences of discrete public actions.

The first approach models influence as a **nonlinear, continuous function of an agent's own conviction**. This idea has roots in both control theory and statistical physics. From

[83] "Social Influence and Opinions". Noah E. Friedkin et al. (1990).
[84] "The Problem of Social Control and Coordination of Complex Systems in Sociology". Noah E. Friedkin. (2015).
[85] "Novel Multidimensional Models of Opinion Dynamics in Social Networks". Sergey E. Parsegov et al. (2017).
[86] "Mixing Beliefs among Interacting Agents". Guillaume Deffuant et al. (2000).
[87] "Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation". R. Hegselmann et al. (2002).
[88] "Opinion Dynamics With Decaying Confidence". Irinel-Constantin Morarescu et al. (2011).
[89] "Dynamics of Opinion Forming in Structurally Balanced Social Networks". Claudio Altafini. (2012).
[90] "Consensus Problems on Networks With Antagonistic Interactions". Claudio Altafini. (2013).
[91] "A Lifting Approach to Models of Opinion Dynamics with Antagonisms". Julien M. Hendrickx. (2014).

the control and dynamical systems perspective, a large body of work, such as that by [92] and [93], uses sigmoidal gain functions to model collective decision-making, often inspired by biological systems like honeybee swarms. This work have been significantly extended in [94–97] to also encompasse antagonistic interactions and understand the polarization behaviors in this situation. In this context, the system is designed to exhibit a pitchfork bifurcation, where a group transitions from a state of indecision (consensus at zero) to a committed decision state. The nonlinearity here represents a social reinforcement mechanism that is requiered for breaking deadlock and ensuring robust consensus on a choice. In parallel, work emerging from computational social science, such as [98, 99], has independently adopted a similar sigmoidal $\tanh(\alpha x_j)$ function. Here, the motivation is to model phenomena like group polarization and the formation of echo chambers. The nonlinearity parameter, α, is interpreted as the "controversialness" of a topic; for high α, even moderate opinions exert non negligible influence, creating a feedback loop that drives opinions to extremes [100]. Though originating from different communities, these models share the core mechanism of a nonlinear activation function to explain the emergence of radicalized, non-consensus states.

The second, distinct approach formalizes a difference between an agent's continuous internal opinion and its **discrete public action**. These models, often referred to as Continuous Opinions and Discrete Actions (CODA), were introduced by [101, 102] and assert that influence is driven not by the internal opinion x_j, but by a quantized representation of it, $q(x_j)$. The model have been expressed in a discrete time dynamical system framework in [67] where polarization was shown to be a robust phenomenon. Analysing the continuous counterpart of this model, [103, 104] showed that this seemingly simple modification has profound dynamical consequences, transforming the system into a set of differential equations with a discontinuous right-hand side. The rigorous analysis of such systems requires sophisticated tools for non-smooth dynamics, often involving solutions in the sense of Krasovskii or Filippov. This line of work, advanced by [105], has demonstrated that the act of quantization itself is a potent source of disagreement. By restricting the information agents can exchange, this mechanism can robustly prevent global consensus, leading to persistent disagreement clusters. More recent work has shown that this framework can even produce highly complex dynamics, including stable limit cycles and Zeno phenomena, where an infinite number of switching events occur in finite time [106, 107].

While historically developing in parallel, these two approaches (one based on smooth nonlinear activation and the other on discontinuous quantized action) are conceptually convergent. Both explain the failure of consensus not by the structure of the network (e.g., negative ties), but by the very nature of the social interaction process. They shift the focus from *who* influences

[92] "Multiagent Decision-Making Dynamics Inspired by Honeybees". Rebecca Gray et al. (2018).
[93] "The Nonlinear Feedback Dynamics of Asymmetric Political Polarization". Naomi Ehrich Leonard et al. (2021).
[94] *A Model-Independent Theory of Consensus and Dissensus Decision Making*. Alessio Franci et al. (2020).
[95] "Breaking Indecision in Multiagent, Multioption Dynamics". Alessio Franci et al. (2023).
[96] "Nonlinear Opinion Dynamics With Tunable Sensitivity". Anastasia Bizyaeva et al. (2023).
[97] "Multi-Topic Belief Formation Through Bifurcations". Anastasia Bizyaeva et al. (2025).
[98] "Modeling Echo Chambers and Polarization Dynamics in Social Networks". Fabian Baumann et al. (2020).
[99] "Emergence of Polarized Ideological Opinions in Multidimensional Topic Spaces". Fabian Baumann et al. (2021).
[100] "Transitions between Polarization and Radicalization in a Echo-Chamber". Łukasz G. Gajewski et al. (2022).
[101] "Continuous Opinions and Discrete Actions in Opinion Dynamics Problems". André C. R. Martins. (2008).
[102] "Discrete Opinion Models as a Limit Case of the CODA Model". André C.R. Martins. (2014).
[103] "Continuous-Time Consensus Dynamics with Quantized All-to-All". Francesca Ceragioli et al. (2015).
[104] "Discontinuities, Generalized Solutions, and (Dis)Agreement in Opinion Dynamics". F. Ceragioli et al. (2018).
[105] "Consensus and Disagreement". Francesca Ceragioli et al. (2018).
[106] "Disagreement, Limit Cycles and Zeno Solutions in COBA". Raoul Prisant et al. (2024).
[107] "A Continuous Opinions Discrete Actions Model on Line Graphs". Francesca Ceragioli et al. (2025).

whom to *how* an agent's internal state is translated into an influential signal. The emerging synthesis, which this thesis aims to advance, recognizes that both nonlinear reinforcement and communicative constraints are critical mechanisms, offering a richer, more nuanced framework for understanding the complex, clustered, and often polarized opinion landscapes we observe in reality.

These four research thrusts, bounded confidence, antagonistic interactions, agent stubbornness, and perception based influence, represent an evolution from the original consensus paradigm. They demonstrate that by carefully introducing psychologically and sociologically plausible mechanisms into the dynamics, multi-agent models can successfully explain not only agreement but also the complex, clustered, and often polarized opinion landscapes we observe in reality.

1.6 Outline of the Manuscript

1.6.1 Contributions

The primary contributions of this manuscript are centered on the development and rigorous analysis of a general framework for nonlinear opinion dynamics, leading to a deeper understanding of synchronization and clustering phenomena in multi-agent systems.

- **A general framework for nonlinear opinion dynamics:** We generalize the Continuous Opinion Discrete Action (CODA) model to the Continuous Opinion Signaled Action (COSA) model, a versatile framework that generalizes existing models by allowing for a large class of nonlinear communication signals. By establishing the forward invariance of the fully synchronized manifold (that is the set where all agents are synchronized), we make the link with the one dimensional dynamics and establish the relation between the equilibria of the system and the fixed points of the interaction function. Moreover, we show that the system is inherently cooperative, which ensures that all (or almost all) trajectories converge to the set of equilibria and precludes complex oscillatory or chaotic behaviors.
- **Global structure of the state space:** We characterize the global behavior of the system by proving that the state space is partitioned into invariant hypercubes whose boundaries are defined by the unstable fixed points of the signal function. This result allows us to localize trajectories and characterize the basins of attraction for synchronized equilibria in the high-dimensional space based solely on the properties of the scalar signal function.
- **A sharp threshold for global synchronization:** We provide a comprehensive characterization of the conditions leading to full network consensus. To achieve this, we introduce a *natural geometry* for the system based on a degree-weighted inner product, which renders the interaction operator self-adjoint and facilitates rigorous Lyapunov analysis. The central result is the derivation of a sharp threshold, $K\lambda_{N-1} < 1$, which elegantly balances the cohesive strength of the network topology (quantified by its spectral properties) against the divisive potential of the signal nonlinearity (quantified by its Lipschitz constant K). This result provides an elegant and verifiable condition for predicting the emergence of global agreement.
- **A structural theory of perfect and robust clustering:** We analyze the structure of disagreement when global consensus fails.
  - We prove that the formation of *perfect* clusters, where agents within a group reach

exact agreement, is not a generic phenomenon but is strictly constrained by structural symmetries in the network topology, namely the existence of a regular partition. This allows for an exact model reduction to a lower-dimensional quotient system. Furthermore, we derive necessary and sufficient conditions for the stability of such clusters, decoupling the analysis into quotient stability (inter-cluster dynamics) and transversal stability (intra-cluster cohesion).

– For networks lacking such perfect symmetry, we develop a novel framework for analyzing *approximated local synchronization*. By employing the control-theoretic tools of Input-to-State Stability (ISS), we quantify the robustness of clusters and their ability to maintain internal coherence against perturbations from the rest of the network.

1.6.2 Chapter contents

– Networked Systems Preliminaries establishes the necessary mathematical preliminaries. It provides a self-contained review of concepts from algebraic graph theory, the theory of continuous-time dynamical systems with a focus on cooperative and monotone systems, and Lyapunov stability theory.
– A non-linear opinion dynamics model: Continuous Opinion Signaled Action (COSA) formally introduces the central theoretical object of the thesis: the Continuous Opinion Signaled Action (COSA) model. This chapter details the model's construction, establishes its core properties such as cooperativity and the invariance of the state space, and provides a complete characterization of the dynamics on the fully synchronized manifold.
– The Full Synchronization in COSA is dedicated to a comprehensive analysis of the full synchronization regime. Building from local stability of synchronized equilibria, it derives conditions for global attractivity and concludes with the derivation of a sharp spectral threshold that guarantees global convergence to consensus for the entire network.
– Partial Synchronization in COSA investigates the phenomena of partial synchronization and clustering that emerge when global consensus is not achieved. It first establishes the necessary structural conditions for the existence of perfect clusters (equitable partitions) and then develops a framework based on Input-to-State Stability (ISS) to analyze the robustness of imperfect clusters.
– Conclusions and Perspectives concludes the thesis by providing a summary of the main contributions and perspectives for future research.

1.6.3 List of publications

Journal papers

[46] Mroué, B., Couthures, A., Lasaulce, S., & Morărescu, I.-C. (2025). On the impact of cross-country imitation on climate change: A game-theoretical analysis. IFAC Journal of Systems and Control, 32, 100309. https://doi.org/10.1016/j.ifacsc.2025.100309

[68] Couthures, A., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (2024). Analysis of an opinion dynamics model coupled with an external environmental dynamics. Chaos, Solitons & Fractals, 189, 115719. https://doi.org/10.1016/j.chaos.2024.115719

[108] Couthures, A., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (2025). Global Synchronization of Multi-Agent Systems With Nonlinear Interactions. IEEE Control Systems Letters, 9, 354–359. https://doi.org/10.1109/LCSYS.2025.3573563

Submitted papers

- Couthures, A., Bainier, G., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (submitted). From global synchronization to robust clustering: multi-agent systems with nonlinear interactions. IEEE Transactions on Automatic Control
- Couthures, A., Mroué, B., Lasaulce, S., & Morărescu, I.-C. (submitted). International Climate Policy as a Potential Game: Strategic Incentives and Path-Dependent Dynamics. Journal of Environmental Economics and Management
- Couthures, A., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (submitted). Link between equitable partitions and local agreements in multi-agent systems with nonlinear interactions. Nonlinear Analysis: Hybrid Systems

Conference papers

[42] Mroué, B., Couthures, A., Lasaulce, S., & Morărescu, I.-C. (2023). Une approche par la théorie des jeux du problème du choix des niveaux nationaux d'émissions de CO2. XXIXème Colloque Francophone de Traitement Du Signal et Des Images, GRETSI 2023. https://gretsi.fr/data/colloque/pdf/2023_mroue1148.pdf

[45] Mroué, B., Couthures, A., Lasaulce, S., & Morărescu, I.-C. (2024). A Quadratic Static Game Model for Assessing the Impact of Climate Change. In H. Cherifi, L. M. Rocha, C. Cherifi, & M. Donduran (Eds.), Complex Networks & Their Applications XII (pp. 383–394). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-53499-7_31

[66] Couthures, A., Mongaillard, T., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (2024). Analysis of a Continuous Opinion and Discrete Action Model Coupled with an External Dynamics. 2024 European Control Conference (ECC), 3392–3397. https://doi.org/10.23919/ECC64448.2024.10591291

[109] Couthures, A., Vogler, P., Beaude, O., Lasaulce, S., & Morărescu, I.-C. (2025). An Optimization Setup of the Decarbonization Problem in the Transportation Sector. In H. Le Cadre, Y. Hayel, B. Tuffin, & T. Chahed (Eds.), Network Games, Artificial Intelligence, Control and Optimization (pp. 131–141). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-78600-6_12

[69] Couthures, A., Bizyaeva, A., Satheeskumar Varma, V., Franci, A., & Morărescu, I.-C. (2025). Bifurcation analysis of an opinion dynamics model coupled with an environmental dynamics arXiv:2504.03419. https://doi.org/10.48550/arXiv.2504.03419

[110] Couthures, A., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (2025). Global synchronization of multi-agent systems with nonlinear interactions arXiv:2503.10205. https://doi.org/10.48550/arXiv.2503.10205

Submitted conference papers

- Couthures, A., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (submitted). Link between equitable partitions and local agreements in multi-agent systems with nonlinear interactions. IFAC World Congress 2026

1.6.4 Other activities

Software

[43] Couthures, A. (2025). AnthonyCouthures/CO2-Game [Jupyter Notebook]. https://github.com/AnthonyCouthures/CO2-Game (Original work published 2022)
Documentation: https://anthonycouthures.github.io/CO2-Game/

[111] Couthures, A. (2024). AnthonyCouthures/Private-Transport-Sector-Optimization-Model [Jupyter Notebook]. https://github.com/AnthonyCouthures/Private-Transport-Sector-Optimization-Model (Original work published 2024)

[112] Couthures, A. (2025). AnthonyCouthures/COSA-Simulator [Python]. https://github.com/AnthonyCouthures/COSA-Simulator (Original work published 2025)

Extended abstracts

- Couthures, A., Mroué, B., Morărescu, I.-C., & Lasaulce, S. (2022). A static game model for analyzing country carbon emissions. PGMO days (Programme Gaspard Monge pour l'Optimisation) 2022
- Couthures, A., Mroué, B., Lasaulce, S., & Morărescu, I.-C. (2023). A Planning Game for Global Decarbonization Strategies. PGMO days (Programme Gaspard Monge pour l'Optimisation) 2023.
- Mongaillard, T., Lasaulce, S., Satheeskumar Varma, V., Couthures, A., & Sabir, E. (2023). Optimal Carbon Pricing Based on Dynamical Climate Model. PGMO days (Programme Gaspard Monge pour l'Optimisation) 2023.
- Couthures, A., Satheeskumar Varma, V., Lasaulce, S., & Morărescu, I.-C. (2024). Analysis of a non-linear opinion dynamics model coupled with an external signal. 7th Conference on Analysis and Control of Nonlinear Dynamics and Chaos, ACNDC 2024
- Couthures, A., Mroué, B., Lasaulce, S., & Morărescu, I.-C. (2024). On the impact of cross-country imitation on climate change: A game-theoretical analysis. French Association of Environmental and Resource Economists, FAERE 2024.

Visit to other laboratories

During the course of this thesis, I had the opportunity to collaborate with external research teams through the following visits:

– **EDF Lab Paris Saclay (Palaiseau, France):**
I spent one month in the R&D OSIRIS department working on the ProspER project. The objective of this visit was to apply an optimization framework to the problem of decarbonizing the private transportation sector. This collaboration led to the development of a specific simulation tool [111] and the publication of the conference paper [109].

– **Institut Montefiore, University of Liège (Liège, Belgium):**
I visited the Department of Electrical Engineering and Computer Science for one week. The discussions focused on theoretical extensions of the coupling between Opinion Dynamics and Environmental dynamics. Specifically, we explored a new theoretical framework to analyze the feedback mechanisms between the vector of influence between the opinion dynamics and the environmental dynamics, as an extension to the analysis [69].

Scientific mediation

- Participation at "Ma Thèse en 180 secondes" 2024 event ("Three minutes Thesis"). This event is a competition where PhD students have to present their thesis in 180 seconds to a non scientific audience. I was Regional finalist. Video available at https://ultv.univ-lorraine.fr/video/17235-ma-these-en-180-secondes-2024-anthony-couthures/.

Research Group

During this thesis I got the opportunity to participate to diverse actions of the Research Group "Modélisation, Analyse et Conduite des Systèmes Complexes (MACS)":

- Presentation of my research during a working group meeting on the topic of "Opinion Dynamics and Environmental Dynamics".

- Presentation of a poster during the "Grand séminaire MACS - Systèmes multi-agents" event. This workshop is organized by the Research Group MACS, https://sma.uphf.fr/. The poster is available at https://hal.science/hal-04715787v1/document.

Chapter 2

Networked Systems Preliminaries

2.1 Introduction

This chapter establishes the foundational concepts and tools essential for modeling and analyzing interconnected dynamical systems. Networked systems, prevalent in various domains from social and biological interactions to engineered infrastructures, exhibit complex collective behaviors that necessitate a systematic analytical approach.

We begin by introducing **Algebraic Graph Theory**, a framework for representing and quantifying network structures. This section defines key graph components such as nodes, edges, and their connectivity properties, and details how these structures can be translated into matrix representations. These matrix forms enable the application of linear algebra for the quantitative analysis of network characteristics.

Following this, we delve into the basics of **Dynamical Systems** theory. As networked systems are inherently dynamic, their states evolve over time governed by interactions between components. We introduce the general form of a continuous-time autonomous dynamical system (and its properties)

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)),$$

and define core concepts such as solutions, trajectories, and equilibrium points. This section also explores specialized classes of systems, including cooperative and monotone systems, which exhibit predictable order-preserving behaviors. This will be important for the rest of the thesis where we will study the behavior of networked systems under various dynamics.

A critical aspect of understanding these dynamic processes is **Stability Theory**. This section outlines various stability notions, from basic boundedness to exponential convergence, providing the means to assess how system states behave in the vicinity of equilibria or other invariant sets. These stability criteria are indispensable for predicting long-term network behavior.

Finally, we present a set of fundamental **Mathematical Tools** that underpin the rigorous development and proofs throughout the thesis. These preliminaries provide a robust analytical foundation, preparing the reader for subsequent chapters where these concepts will be applied to study phenomena such as synchronization, consensus, and the impact of external influences on multi-agent systems.

2.2 Algebraic Graph Theory

This section introduces the basic concepts of graph theory and algebraic graph theory, which are essential for modeling and analyzing interconnected dynamical systems. Graphs provide a language to describe relationships and interactions within a network, while algebraic graph theory equips us with tools to analyze these structures numerically using matrices. The concepts and results presented here will establish a solid foundation for understanding network behavior in various dynamical models that will be required in the rest of the thesis.

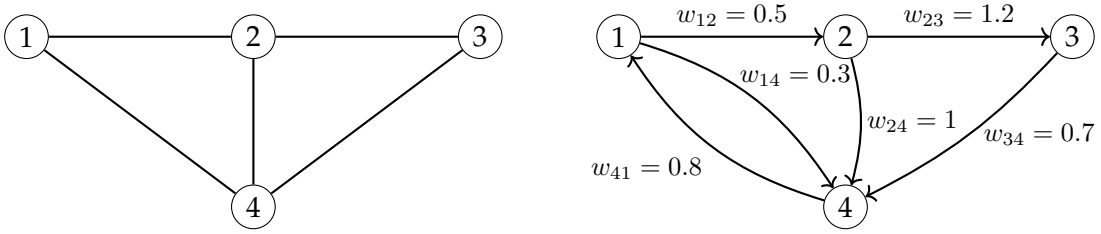
2.2.1 Basic Graph Concepts

We begin by defining the core components of a graph, which serves as our abstract representation of a network. A graph consists of a set of vertices and a set of connections between them, which can be either undirected or directed, and optionally assigned weights. The definitions and results presented here are taken from [113].

Definition 2.2.1 (Graph)**.** *A **graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is formally defined as a pair consisting of a finite non-empty set of **vertices** (or **nodes**) $\mathcal{V}$ and a set of **edges** $\mathcal{E}$.*

- *An **undirected graph** has edges as unordered pairs of distinct vertices $\{i, j\}$.*
- *A **directed graph** (or **digraph**) has edges as ordered pairs of distinct vertices (i, j), indicating a connection from i to j.*
- *A **weighted graph** assigns a real number $w_{ij} > 0$ to each edge (i, j) or $\{i, j\}$, representing the strength or capacity of the connection, and $w_{ij} = 0$ if the edge (i, j) does not exist. When the weight is not specified, the graph is **unweighted** and we assume it to be 1.*

We will assume that $\mathcal{V} = \{1, 2, \ldots, N\}$ for some $N \in \mathbb{N}$ with $N \geq 2$.



(a) A simple undirected, unweighted graph $\mathcal{G}_1$ with edges set $\mathcal{E} = \{\{1, 2\}, \{2, 3\}, \{1, 4\}, \{2, 4\}, \{3, 4\}\}$.

(b) A simple weighted directed graph (digraph) $\mathcal{G}_2$ with edges set $\mathcal{E} = \{(1, 2), (2, 3), (1, 4), (2, 4), (3, 4), (4, 1)\}$ and weights w_{ij}.

Figure 2.1: A simple undirected, unweighted graph $\mathcal{G}_1$ (left) and a simple weighted directed graph $\mathcal{G}_2$ (right). The vertices are labeled as $\mathcal{V} = \{1, 2, 3, 4\}$.

Throughout this thesis, we will predominantly refer to undirected, unweighted, and simple graphs unless otherwise specified.

Definition 2.2.2 (Node and Edge)**.** *A **node** (or **vertex**) is an element $i \in \mathcal{V}$ representing an entity in the network.*

[113] *Algebraic Graph Theory*. Chris Godsil et al. (2001).

- *For an* undirected graph, *an* ***edge*** $\{i, j\}$ *represents a symmetric connection between node* i *and node* j*. If* $\{i, j\} \in \mathcal{E}$*, then* $(i, j) \in \mathcal{E}$ *and* $(j, i) \in \mathcal{E}$*.*
- *For a* directed graph, *an* ***edge*** (i, j) *represents an ordered connection from node* i *to node* j*. Node* i *is the* ***tail*** *and node* j *is the* ***head****.*

We say that the graph is ***simple*** *if it has no loops (edges that connect a vertex to itself) or multiple edges (edges with the same tail and head). i.e.,* $\mathcal{E}$ *is a set of unordered pairs of distinct vertices.*

Definition 2.2.3 (Neighborhood and Degree)**.** *To each node* $i \in \mathcal{V}$*, we associate the notion of* ***neighborhood*** *and* ***degree****.*

- *In an* undirected graph, *the* ***neighborhood*** $\mathcal{N}_i$ *of vertex* i *is*

$$\mathcal{N}_i = \{j \in \mathcal{V} \mid \{i, j\} \in \mathcal{E}\}.$$

  The ***degree*** d_i *is* $|\mathcal{N}_i|$ *since every edge has weight* 1*. For a* weighted undirected graph, *the* ***weighted degree*** *of* i *is*

$$d_i = \sum_{j \in \mathcal{N}_i} w_{ij}.$$

- *In a* directed graph *(digraph):*
  - *The* ***out-neighborhood*** $\mathcal{N}_i^{\text{out}}$ *of* i *is* $\mathcal{N}_i^{\text{out}} = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}$*. The* ***out-degree*** d_i^{out} *is* $d_i^{\text{out}} = \sum_{j \in \mathcal{N}_i^{\text{out}}} w_{ij}$*.*
  - *The* ***in-neighborhood*** $\mathcal{N}_i^{\text{in}}$ *of* i *is* $\mathcal{N}_i^{\text{in}} = \{j \in \mathcal{V} \mid (j, i) \in \mathcal{E}\}$*. The* ***in-degree*** d_i^{in} *is* $d_i^{\text{in}} = \sum_{j \in \mathcal{N}_i^{\text{in}}} w_{ji}$*.*

Let us now compute the neighborhood and degree for the graphs in Figure 2.1a and Figure 2.1b as an example.

Example 2.2.1. *For the undirected graph in Figure 2.1a:* $\mathcal{N}_1 = \{2, 4\}$ *and* $\mathcal{N}_2 = \{1, 3, 4\}$ *with* $d_1 = 2$ *and* $d_2 = 3$ *and so on.*

For the weighted directed graph in Figure 2.1b: $\mathcal{N}_1^{\text{out}} = \{2, 4\}$ *and* $\mathcal{N}_1^{\text{in}} = \{1\}$ *with* $d_1^{\text{out}} = 0.8$ *and* $d_1^{\text{in}} = 0.8$*. Similarly,* $\mathcal{N}_2^{\text{out}} = \{3, 4\}$ *and* $\mathcal{N}_2^{\text{in}} = \{1\}$ *with* $d_2^{\text{out}} = 2.2$ *and* $d_2^{\text{in}} = 0.5$ *and so on.*

In graph theory, sequence of vertices connected by edges are central to the study of graph properties. We will now define the notion of a path linking two vertices.

Definition 2.2.4 (Path)**.** *A* ***path*** *(or* ***walk****) in an graph is a ordered sequence of vertices/edges* $i_1, i_2, \ldots, i_{p+1}$ *such that* $\{i_k, i_{k+1}\} \in \mathcal{E}$ *for all* $k \in \{1, \ldots, p\}$*. All edges must follow their specified direction.*

Having defined the notion of a path, we can now define the notion of connectedness in the graph.

Definition 2.2.5 (Connectedness)**.** *In a graph* $\mathcal{G}$*, two vertices* i *and* j *are* ***connected*** *if there exists a path between them. Such notion of connectedness extends to the entire graph:*

- *An undirected graph is* ***connected*** *if all pairs of vertices are connected.*
- *In a* directed graph *(digraph), we refine the notion of connectedness:*
  - *A digraph is* ***weakly connected*** *if its underlying undirected graph (ignoring edge directions) is connected.*
  - *A digraph is* ***strongly connected*** *if for every ordered pair of vertices* (i, j)*, there exists*



(a) A disconnected graph $\mathcal{G}$ with two connected components: $A = \{1, 2, 3\}$ and $B = \{4, 5\}$.

(b) A subgraph $\mathcal{G}' \subset \mathcal{G}$ with two connected components: $A' = \{1, 2\}$ and $B' = \{4\}$.

Figure 2.2: Illustration of the notion of a connected component. The graph is disconnected with two connected components: $\{1, 2, 3\}$ and $\{4, 5\}$.

a directed path from i to j.

*A **connected component** is a maximal connected subgraph.*

Figure 2.2a illustrates an undirected graph with two distinct connected components.

It is then intuitive to define the notion of a subgraph, which is a subset of the vertices and edges of a graph.

Definition 2.2.6 (Subgraph and Induced Subgraph)**.** *A **subgraph** $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a graph such that $\mathcal{V}' \subseteq \mathcal{V}$ and $\mathcal{E}' \subseteq \mathcal{E}$, with the additional condition that all edges in $\mathcal{E}'$ must connect vertices within $\mathcal{V}'$.*

*An **induced subgraph** is a subgraph formed by choosing a subset of vertices $\mathcal{V}' \subseteq \mathcal{V}$ and $\mathcal{E}' = \mathcal{E} \cap (\mathcal{V}' \times \mathcal{V}')$. i.e., the nodes in $\mathcal{V}'$ and the edges that connect two vertices in $\mathcal{V}'$.*

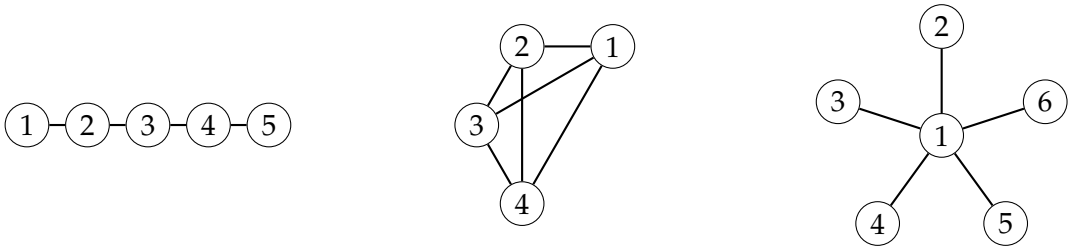
Figure 2.2 illustrates the notion of a subgraph as a subset of the vertices and edges of a graph presented in Figure 2.2a. Note that a subgraph of $\mathcal{G}$ is not necessarily a connected component of $\mathcal{G}$ since it is not connected.

2.2.2 Canonical Graphs

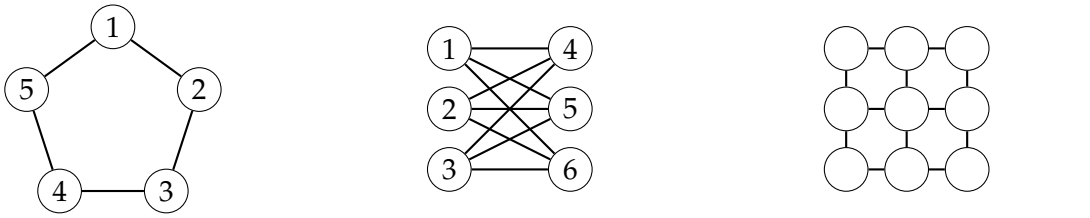
Canonical graphs are mathematical objects that appear frequently in graph theory and network analysis. They serve as building blocks and illustrative examples for various graph properties.

Definition 2.2.7. *We define the following canonical graphs:*

- *A **path graph** P_N is a graph with N vertices and edges connecting vertices in order.*
- *A **complete graph** K_N is a graph with N vertices and edges connecting every pair of vertices.*
- *A **star graph** S_N is a graph with N vertices and one central vertex connected to all other vertices.*
- *A **cycle graph** C_N is a graph with N vertices and edges connecting vertices in a cycle.*
- *A **bipartite graph** $B_{M,N}$ is a graph with two sets of vertices, $\mathcal{V}_1$ with M vertices and $\mathcal{V}_2$ with N vertices, such that every edge in $\mathcal{E}$ connects a vertex in $\mathcal{V}_1$ to a vertex in $\mathcal{V}_2$.*
- *A **complete bipartite graph** $K_{M,N}$ is a graph with two sets of vertices, $\mathcal{V}_1$ with M vertices and $\mathcal{V}_2$ with N vertices, such that every vertex in $\mathcal{V}_1$ is connected to every vertex in $\mathcal{V}_2$, and there are no edges within $\mathcal{V}_1$ or $\mathcal{V}_2$.*
- *A **2D lattice** (or **grid graph**) $G_{M,N}$ is a graph with $M \times N$ vertices arranged in a grid,*



(a) Path Graph (P_5) (b) Complete Graph (K_4) (c) Star Graph (S_6)



(d) Cycle Graph (C_5) (e) Complete Bipartite Graph ($K_{3,3}$) (f) 2D Lattice (Grid Graph, 3×3)

Figure 2.3: Illustrations of Canonical Graph Structures.

where each vertex is connected to its four nearest neighbors (up, down, left, right) except for the boundary vertices.

The canonical graphs are illustrated in Figure 2.3. Those graphs are useful for illustrating the properties of graphs and for testing the correctness of algorithms and will be used throughout the thesis for illustration purposes.

2.2.3 Graph Representations: Matrices

To perform quantitative analysis of network structures and dynamics, we represent graphs using matrices. These matrices encode the connectivity and other properties of the graph in a mathematical form, enabling the application of linear algebra tools. The definitions and results presented here are inspired by [113, 114].

Definition 2.2.8 (Adjacency Matrix). *The **adjacency matrix** $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ for a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $N = |\mathcal{V}|$ vertices specifies the connections between nodes. Its entry a_{ij} represents the weight of the edge from node i to node j.*

- *For an* unweighted graph, $a_{ij} = 1$ *if an edge exists, and* 0 *otherwise.*
- *For a* weighted graph, $a_{ij} = w_{ij}$ *if an edge exists, and* 0 *otherwise.*
- *For an* undirected graph, $\boldsymbol{A}$ *is symmetric* ($a_{ij} = a_{ji}$).
- *For a* directed graph *(digraph),* $\boldsymbol{A}$ *is generally asymmetric.*

Let us continue to illustrate the notion of adjacency matrix on the graphs presented in Figure 2.1a and Figure 2.1b.

[114] *Lectures on Network Systems*. Francesco Bullo. (2018).

Example 2.2.2. *For the undirected graph $\mathcal{G}_1$ in Figure 2.1a and the weighted directed graph $\mathcal{G}_2$ in Figure 2.1b, the adjacency matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are respectively:*

$$\boldsymbol{A}_1 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \textit{and} \quad \boldsymbol{A}_2 = \begin{pmatrix} 0 & 0 & 0 & 0.8 \\ 0.5 & 0 & 0 & 0 \\ 0 & 1.2 & 0 & 0 \\ 0.3 & 1 & 0.7 & 0 \end{pmatrix}.$$

Definition 2.2.9. *The **degree matrix** $\boldsymbol{D} \in \mathbb{R}^{N \times N}$ of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a diagonal matrix where each diagonal entry d_{ii} is the degree of vertex i. i.e., $d_{ii} = \sum_{j=1}^{N} a_{ij} = |\mathcal{N}_i|$.*

Continue with the example of the graphs presented in Figure 2.1a and Figure 2.1b, we can now define their respective degree matrices.

Example 2.2.3. *For the graph in Figure 2.1a, the degree matrix $\boldsymbol{D}_1$ of $\mathcal{G}_1$ and the degree matrices $\boldsymbol{D}^{\text{out}}$ and $\boldsymbol{D}^{\text{in}}$ of $\mathcal{G}_2$ are:*

$$\boldsymbol{D}_1 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \textit{ and } \boldsymbol{D}_2^{\text{out}} = \begin{pmatrix} 0.8 & 0 & 0 & 0 \\ 0 & 2.2 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.8 \end{pmatrix} \textit{ and } \boldsymbol{D}_2^{\text{in}} = \begin{pmatrix} 0.8 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1.2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

One can note define the family of graphs for which the degree matrix is a diagonal matrix.

Definition 2.2.10. *Let $d \in \mathbb{N}$. A undirected graph is called d-**regular** if all its vertices have the same degree d. i.e., $\forall i \in \mathcal{V}, d_i = d$ or equivalently $\boldsymbol{D} = d\boldsymbol{I}$.*

Being equipped with the notion of adjacency matrix and degree matrix, we can now define the *Laplacian matrix* that will be used throughout the thesis.

Definition 2.2.11. *The **Laplacian matrix** $\boldsymbol{L} \in \mathbb{R}^{N \times N}$ of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined as the difference between the degree matrix and the adjacency matrix:*

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}.$$

With the convention that $\boldsymbol{D} = \boldsymbol{D}^{\text{in}}$ for directed graphs.

For the graph in Figure 2.1a, the Laplacian matrices $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are:

$$\boldsymbol{L}_1 = \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} \quad \text{and} \quad \boldsymbol{L}_2 = \begin{pmatrix} 0.8 & -0.5 & 0 & -0.3 \\ 0 & 2.2 & -1.2 & -1 \\ 0 & 0 & 0.7 & -0.7 \\ -0.8 & 0 & 0 & 0.8 \end{pmatrix}.$$

The Laplacian matrix is symmetric for undirected graphs, and its rows (and columns) sum to zero, which implies that $\boldsymbol{L}\mathbf{1} = \mathbf{0}$, where $\mathbf{1}$ is the vector of all ones. This property indicates that 0 is always an eigenvalue of $\boldsymbol{L}$, with $\mathbf{1}$ as its corresponding eigenvector.

Normalized Graph Matrices

For certain network dynamics, particularly those involving averaging or distributed updates, it is often useful to work with normalized versions of these matrices. Normalization helps to standardize the influence or activity of each node, regardless of its raw degree.

Definition 2.2.12. *The **left-normalized adjacency matrix** (or **random walk matrix**) $\boldsymbol{A}_{\mathrm{rw}} \in \mathbb{R}^{N\times N}$ is defined as $\boldsymbol{A}_{\mathrm{rw}} = \boldsymbol{D}^{-1}\boldsymbol{A}$, with the convention that $\boldsymbol{D} = \boldsymbol{D}^{\mathrm{in}}$ for directed graphs.*

*The **left-normalized Laplacian matrix** $\boldsymbol{L}_{\mathrm{rw}} \in \mathbb{R}^{N\times N}$ is defined as $\boldsymbol{L}_{\mathrm{rw}} = \boldsymbol{D}^{-1}\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{A}$.*

Example 2.2.4. *For the graph in Figure 2.1a, the left-normalized adjacency matrices $\boldsymbol{A}_{\mathrm{rw}}$ and $\boldsymbol{L}_{\mathrm{rw}}$ are:*

$$\boldsymbol{A}_{\mathrm{rw}} = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix} \quad \textit{and} \quad \boldsymbol{L}_{\mathrm{rw}} = \begin{pmatrix} 1 & -1/2 & 0 & -1/2 \\ -1/3 & 1 & -1/3 & -1/3 \\ 0 & -1/2 & 1 & -1/2 \\ -1/3 & -1/3 & -1/3 & 1 \end{pmatrix}.$$

The left-normalized adjacency matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$ is a **row-stochastic matrix**. This means all its entries are non-negative and each row sum equals 1. This property is especially important in many network dynamics because:

- It inherently constrains the state variables to a bounded domain, provided the update function itself maps into that domain.
- It effectively equalizes the total "outgoing influence" from each agent. In dynamics where agents aggregate information from neighbors, normalization by degree ensures that agents with many connections do not inherently dominate over less connected agents simply by virtue of their degree.

Spectral Properties of Graph Matrices

The matrix representations allow us to leverage powerful results from linear algebra to rigorously analyze graph properties and network behavior.

Perron-Frobenius Theorem. *[114, Theorem 2.12] Let $\boldsymbol{M} \in \mathbb{R}^{N\times N}$ be a non-negative matrix (i.e., all entries $m_{ij} \geq 0$).*

There exists a real eigenvalue λ_P (called the Perron root or spectral radius $\rho(\boldsymbol{M})$) such that $\lambda_P \geq |\mu|$ for all other eigenvalues μ of $\boldsymbol{M}$.

(ii) There exists a non-negative right eigenvector $\boldsymbol{v}_P \geq \boldsymbol{0}$ (i.e., $v_i \geq 0$ for all i) and a non-negative left eigenvector $\boldsymbol{w}_P \geq \boldsymbol{0}$ corresponding to λ_P.

*(iii) If $\boldsymbol{M}$ is **irreducible** (meaning its associated directed graph is strongly connected), then λ_P is strictly positive, simple (multiplicity 1), and its corresponding eigenvectors $\boldsymbol{v}_P \gg \boldsymbol{0}$ (i.e., $v_i > 0$ for all i) and $\boldsymbol{w}_P \gg \boldsymbol{0}$ are unique (up to scaling).*

***Application for normalized adjacency matrices:** For a connected, undirected graph $\mathcal{G}$, its left-normalized adjacency matrix $\boldsymbol{A}_{\mathrm{rw}} = \boldsymbol{D}^{-1}\boldsymbol{A}$ is irreducible (since $\mathcal{G}$ is connected, any node can reach any other node). Furthermore, $\boldsymbol{A}_{\mathrm{rw}}$ is row-stochastic. According to the Perron-Frobenius Theorem:*

The largest eigenvalue (Perron root) is $\lambda_P = 1$. This is because $\boldsymbol{A}_{\mathrm{rw}}\boldsymbol{1} = \boldsymbol{1}$ (sum of each row is 1), so $\boldsymbol{1}$ is a right eigenvector with eigenvalue 1.

- *This eigenvalue $\lambda_P = 1$ is simple, and all other eigenvalues μ of $\boldsymbol{A}_{\mathrm{rw}}$ satisfy $\mu < 1$ and*

$|\mu| \leq 1$.

- *There exists a unique (up to scaling) positive left eigenvector $\boldsymbol{w} \gg \mathbf{0}$ corresponding to $\lambda_P = 1$, such that $\boldsymbol{w}^\top \boldsymbol{A}_{\mathrm{rw}} = \boldsymbol{w}^\top$. This vector $\boldsymbol{w}$ represents the stationary distribution of a random walk on the graph.*

These properties are necessary for understanding the convergence of many network dynamics to a consensus or equilibrium state.

[114] *Lectures on Network Systems*. Francesco Bullo. (2018).

We now state some properties of the eigenvalues of symmetric graph matrices.

Proposition 2.2.13 (Eigenvalues of Symmetric Graph Matrices)**.** *For an* undirected graph*, both the adjacency matrix $\boldsymbol{A}$ and the Laplacian matrix $\boldsymbol{L}$ are symmetric. Therefore, all their eigenvalues are real. Additionally, for an* undirected graph*, the Laplacian matrix $\boldsymbol{L}$ is* ***positive semi-definite****, meaning all its eigenvalues are non-negative.*

Proof. This is a direct consequence of matrix theory of symmetric matrices [115].

For an undirected graph, the Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$. For any vector $\boldsymbol{x} \in \mathbb{R}^N$, the quadratic form $\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}$ can be written as:

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \sum_{i=1}^{N} d_i x_i^2 - \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij} x_i x_j = \sum_{i=1}^{N}\sum_{j\in\mathcal{N}_i} w_{ij} x_i^2 - \sum_{i=1}^{N}\sum_{j\in\mathcal{N}_i} w_{ij} x_i x_j,$$

since $d_i = \sum_{j\in\mathcal{N}_i} w_{ij}$ and $a_{ij} = w_{ij}$ with the convention that $w_{ij} = 1$ if $\{i,j\} \in \mathcal{E}$. Moreover, since the graph is undirected, $a_{ij} = a_{ji}$, the second sum can be rewritten as:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij} x_i x_j = 2 \sum_{\{i,j\}\in\mathcal{E}} w_{ij} x_i x_j = 2 \sum_{i=1}^{N}\sum_{j\in\mathcal{N}_i} w_{ij} x_i x_j.$$

Thus, because each edge $\{i,j\}$ appears twice in the sum (once for i and once for j), we can rewrite:

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \sum_{\{i,j\}\in\mathcal{E}} w_{ij}(x_i^2 + x_j^2 - 2x_i x_j) = \sum_{\{i,j\}\in\mathcal{E}} w_{ij}(x_i - x_j)^2$$

Finally, since $w_{ij} \geq 0$ and $(x_i - x_j)^2 \geq 0$, it follows that $\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^N$. This proves that $\boldsymbol{L}$ is positive semi-definite, and thus all its eigenvalues are non-negative. □

One can formulate the following proposition regarding the eigenvalues of bipartite graphs.

Proposition 2.2.14 (Eigenvalues of Bipartite Graphs)**.** *Let $\mathcal{G}$ be a bipartite graph. If λ is an eigenvalue of its adjacency matrix $\boldsymbol{A}$, then $-\lambda$ is also an eigenvalue of $\boldsymbol{A}$. The same holds for the random walk adjacency matrix $\boldsymbol{A}_{\mathrm{rw}}$.*

Proof. Since $\mathcal{G}$ is a bipartite graph, its vertices can be partitioned into two sets, V_1 and V_2, such that all edges connect a vertex in V_1 to one in V_2. This induces a block structure on the adjacency matrix $\boldsymbol{A}$ and the diagonal degree matrix $\boldsymbol{D}$:

$$\boldsymbol{A} = \begin{pmatrix} \mathbf{0} & \boldsymbol{B} \\ \boldsymbol{B}^\top & \mathbf{0} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_2 \end{pmatrix}.$$

Let λ be an eigenvalue of $\boldsymbol{A}$ with a corresponding eigenvector $\boldsymbol{v}$, partitioned as $\boldsymbol{v} = (\boldsymbol{v}_1, \boldsymbol{v}_2)^\top$. The eigenvalue equation $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ yields the system:

$$\boldsymbol{B}\boldsymbol{v}_2 = \lambda\boldsymbol{v}_1 \quad \text{and} \quad \boldsymbol{B}^\top\boldsymbol{v}_1 = \lambda\boldsymbol{v}_2.$$

Now, consider the vector $\boldsymbol{v}' = (\boldsymbol{v}_1, -\boldsymbol{v}_2)^\top$. Applying $\boldsymbol{A}$ to $\boldsymbol{v}'$ gives:

$$\boldsymbol{A}\boldsymbol{v}' = \begin{pmatrix} -\boldsymbol{B}\boldsymbol{v}_2 \\ \boldsymbol{B}^\top\boldsymbol{v}_1 \end{pmatrix} = \begin{pmatrix} -\lambda\boldsymbol{v}_1 \\ \lambda\boldsymbol{v}_2 \end{pmatrix} = -\lambda \begin{pmatrix} \boldsymbol{v}_1 \\ -\boldsymbol{v}_2 \end{pmatrix} = -\lambda\boldsymbol{v}'.$$

Thus, $-\lambda$ is also an eigenvalue of $\boldsymbol{A}$ with eigenvector $\boldsymbol{v}'$.

The same logic applies to the random walk matrix $\boldsymbol{A}_{\text{rw}} = \boldsymbol{D}^{-1}\boldsymbol{A}$. Let λ be an eigenvalue of $\boldsymbol{A}_{\text{rw}}$ with eigenvector $\boldsymbol{v}$. The eigenvalue equation $\boldsymbol{A}_{\text{rw}}\boldsymbol{v} = \lambda\boldsymbol{v}$ is equivalent to $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{D}\boldsymbol{v}$. In block form, this is:

$$\boldsymbol{B}\boldsymbol{v}_2 = \lambda\boldsymbol{D}_1\boldsymbol{v}_1 \quad \text{and} \quad \boldsymbol{B}^\top\boldsymbol{v}_1 = \lambda\boldsymbol{D}_2\boldsymbol{v}_2.$$

Using the same transformed vector $\boldsymbol{v}' = (\boldsymbol{v}_1, -\boldsymbol{v}_2)^\top$, we find:

$$\boldsymbol{A}\boldsymbol{v}' = \begin{pmatrix} -\boldsymbol{B}\boldsymbol{v}_2 \\ \boldsymbol{B}^\top\boldsymbol{v}_1 \end{pmatrix} = \begin{pmatrix} -\lambda\boldsymbol{D}_1\boldsymbol{v}_1 \\ \lambda\boldsymbol{D}_2\boldsymbol{v}_2 \end{pmatrix} = -\lambda \begin{pmatrix} \boldsymbol{D}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1 \\ -\boldsymbol{v}_2 \end{pmatrix} = -\lambda\boldsymbol{D}\boldsymbol{v}'.$$

Pre-multiplying by $\boldsymbol{D}^{-1}$ gives $\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{v}' = -\lambda\boldsymbol{v}'$, which shows that $-\lambda$ is an eigenvalue of $\boldsymbol{A}_{\text{rw}}$. □

For the normalized adjacency matrices, we can provide the following relationships between the eigenvalues of the normalized adjacency matrix and the Laplacian matrix.

Proposition 2.2.15 (Eigenvalue Relationships for Normalized Matrices)**.** *Let $\mathcal{G}$ be a graph with adjacency matrix $\boldsymbol{A}$ and degree matrix $\boldsymbol{D}$.*

(i) If λ is an eigenvalue of $\boldsymbol{A}_{\text{rw}} = \boldsymbol{D}^{-1}\boldsymbol{A}$, then $1 - \lambda$ is an eigenvalue of $\boldsymbol{L}_{\text{rw}} = \boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{A}$, and vice versa.

(ii) For a regular undirected graph, *the eigenvalues of $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ are related to the eigenvalues of $\boldsymbol{A}$ by $\lambda_L = d - \lambda_A$.*

Proof. (i) Let λ be an eigenvalue of $\boldsymbol{A}_{\text{rw}}$ with corresponding eigenvector $\boldsymbol{v}$. By definition, $\boldsymbol{A}_{\text{rw}}\boldsymbol{v} = \lambda\boldsymbol{v}$. Now consider the action of $\boldsymbol{L}_{\text{rw}}$ on $\boldsymbol{v}$:

$$\boldsymbol{L}_{\text{rw}}\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{A}_{\text{rw}})\boldsymbol{v} = \boldsymbol{I}\boldsymbol{v} - \boldsymbol{A}_{\text{rw}}\boldsymbol{v} = \boldsymbol{v} - \lambda\boldsymbol{v} = (1 - \lambda)\boldsymbol{v}.$$

Thus, if λ is an eigenvalue of $\boldsymbol{A}_{\text{rw}}$, then $1 - \lambda$ is an eigenvalue of $\boldsymbol{L}_{\text{rw}}$ with the same eigenvector $\boldsymbol{v}$. The converse is proven similarly.

(ii) For a regular undirected graph, every node has the same degree d. Thus, the degree matrix is $\boldsymbol{D} = d\boldsymbol{I}$. The Laplacian matrix is then $\boldsymbol{L} = d\boldsymbol{I} - \boldsymbol{A}$. Let λ_A be an eigenvalue of $\boldsymbol{A}$ with corresponding eigenvector $\boldsymbol{v}$. So, $\boldsymbol{A}\boldsymbol{v} = \lambda_A\boldsymbol{v}$. Now consider the action of $\boldsymbol{L}$ on $\boldsymbol{v}$:

$$\boldsymbol{L}\boldsymbol{v} = (d\boldsymbol{I} - \boldsymbol{A})\boldsymbol{v} = d\boldsymbol{I}\boldsymbol{v} - \boldsymbol{A}\boldsymbol{v} = d\boldsymbol{v} - \lambda_A\boldsymbol{v} = (d - \lambda_A)\boldsymbol{v}.$$

Thus, if λ_A is an eigenvalue of $\boldsymbol{A}$, then $d - \lambda_A$ is an eigenvalue of $\boldsymbol{L}$ with the same eigenvector $\boldsymbol{v}$. □

Proposition 2.2.16 (Similarity and Weighted Self-Adjointness)**.** *For an undirected connected graph $\mathcal{G}$ with degree matrix $\boldsymbol{D}$, the left-normalized adjacency matrix $\boldsymbol{A}_{\mathrm{rw}} = \boldsymbol{D}^{-1}\boldsymbol{A}$ and the symmetric normalized adjacency matrix $\boldsymbol{A}_{\mathrm{sym}} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ are similar and thus have the same set of eigenvalues, denoted by $\lambda_1, \dots, \lambda_N$. Moreover, $\boldsymbol{L}_{\mathrm{rw}}$ and $\boldsymbol{L}_{\mathrm{sym}}$ share the same set of eigenvalues, denoted by $\mu_1, \dots, \mu_N$.*

Consequently:

(i) All eigenvalues of $\boldsymbol{A}_{\mathrm{rw}}$ and $\boldsymbol{L}_{\mathrm{rw}}$ are real.

(ii) $\boldsymbol{A}_{\mathrm{rw}}$ is self-adjoint with respect to the weighted inner product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{D}} := \boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{y}. \tag{2.1}$$

(iii) $\boldsymbol{A}_{\mathrm{rw}}$ is diagonalizable over $\mathbb{R}$, and its eigenvectors can be chosen to form a $\boldsymbol{D}$-orthonormal basis, i.e.

$$\boldsymbol{v}_i^\top \boldsymbol{D}\, \boldsymbol{v}_j = \delta_{ij}. \tag{2.2}$$

Proof. Let $\boldsymbol{S} = \boldsymbol{D}^{1/2}$. Then $\boldsymbol{S}$ is invertible (since $\mathcal{G}$ is connected, all degrees $d_i > 0$), and

$$\boldsymbol{A}_{\mathrm{sym}} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2} = \boldsymbol{S}\, \boldsymbol{A}_{\mathrm{rw}}\, \boldsymbol{S}^{-1}. \tag{2.3}$$

Hence $\boldsymbol{A}_{\mathrm{rw}}$ and $\boldsymbol{A}_{\mathrm{sym}}$ are similar and share the same eigenvalues.

(i) Since $\boldsymbol{A}_{\mathrm{sym}}$ is symmetric, all its eigenvalues are real. The same reasoning applies to $\boldsymbol{L}_{\mathrm{rw}}$ and $\boldsymbol{L}_{\mathrm{sym}}$ via Proposition 2.2.15.

(ii) For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$,

$$\langle \boldsymbol{x}, \boldsymbol{A}_{\mathrm{rw}}\boldsymbol{y} \rangle_{\boldsymbol{D}} = \boldsymbol{x}^\top \boldsymbol{D}\, \boldsymbol{D}^{-1}\boldsymbol{A}\, \boldsymbol{y} = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x} = \langle \boldsymbol{A}_{\mathrm{rw}}\boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{D}}, \tag{2.4}$$

where we used the symmetry of $\boldsymbol{A}$ for an undirected graph. Thus $\boldsymbol{A}_{\mathrm{rw}}$ is self-adjoint in $\langle \cdot, \cdot \rangle_{\boldsymbol{D}}$.

(iii) It follows from the spectral theorem for finite-dimensional self-adjoint operators. □

Proposition 2.2.17 (Properties of the Symmetric Normalized Laplacian)**.** *For an undirected graph $\mathcal{G}$, the **symmetric normalized Laplacian** is defined as $\boldsymbol{L}_{\mathrm{sym}} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$.*

(i) $\boldsymbol{L}_{\mathrm{sym}}$ is a real symmetric matrix, so all its eigenvalues are real.

(ii) $\boldsymbol{L}_{\mathrm{sym}}$ is positive semi-definite, meaning all its eigenvalues are non-negative.

(iii) The multiplicity of the eigenvalue 0 for $\boldsymbol{L}_{\mathrm{sym}}$ is equal to the number of connected components in $\mathcal{G}$.

Proof. (i)–(ii) The proof follows from the one of Proposition 2.2.13.

(iii) Let $\mathcal{E}$ denote the set of edges in $\mathcal{G}$. We examine the null space of $\boldsymbol{L}_{\mathrm{sym}}$. Let $\boldsymbol{x}$ be an eigenvector such that $\boldsymbol{L}_{\mathrm{sym}}\boldsymbol{x} = \boldsymbol{0}$, which implies $\boldsymbol{x}^\top \boldsymbol{L}_{\mathrm{sym}}\boldsymbol{x} = 0$.

Recall that the unnormalized Laplacian is $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, which allows us to write $\boldsymbol{L}_{\mathrm{sym}} = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}$. By defining $\boldsymbol{y} = \boldsymbol{D}^{-1/2}\boldsymbol{x}$, we can rewrite the quadratic form as:

$$\boldsymbol{x}^\top \boldsymbol{L}_{\mathrm{sym}}\boldsymbol{x} = \boldsymbol{y}^\top \boldsymbol{L}\boldsymbol{y} = \sum_{\{i,j\} \in \mathcal{E}} (y_i - y_j)^2 = 0.$$

Because the sum consists of non-negative terms, we must have $y_i = y_j$ for every pair of adjacent vertices $(i, j) \in \mathcal{E}$. Consequently, $\boldsymbol{y}$ must be constant on each connected component of $\mathcal{G}$.

Suppose $\mathcal{G}$ has k connected components $\mathcal{C}_1, \ldots, \mathcal{C}_k$. Since $\boldsymbol{y}$ is piecewise constant, it can be written as a linear combination of the k indicator vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, where the i-th entry of $\boldsymbol{y}_m$ is 1 if $i \in \mathcal{C}_m$ and 0 otherwise.

Transforming back to $\boldsymbol{x}$, any eigenvector in the null space of $\boldsymbol{L}_{\text{sym}}$ must be a linear combination of the vectors $\boldsymbol{x}_m = \boldsymbol{D}^{1/2}\boldsymbol{y}_m$. Because the connected components are disjoint, these k vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ have disjoint supports and are therefore linearly independent. Thus, they form a basis of the eigenspace associated with $\lambda = 0$, proving that the multiplicity of the eigenvalue 0 is exactly k. □

Spectral Gap and Algebraic Connectivity

The eigenvalues of graph matrices provide important insights into the structural properties of a network, particularly its connectivity. The concept of the *spectral gap* or *algebraic connectivity* quantifies how "well-connected" a graph is.

Definition 2.2.18. *For a connected undirected graph $\mathcal{G}$ with Laplacian matrix $\boldsymbol{L}$, let its eigenvalues be ordered $0 = \mu_1 \leq \mu_2 \leq \ldots \leq \mu_N$.*

- *The **algebraic connectivity** (or Fiedler eigenvalue) is μ_2, the second smallest eigenvalue of $\boldsymbol{L}$.*
- *The **spectral gap** of $\boldsymbol{L}$ is $\mu_2 - \mu_1 = \mu_2$.*

A larger spectral gap implies stronger connectivity.

For a normalized adjacency matrix $\boldsymbol{A}_{\text{rw}}$ of a connected graph, we note its eigenvalues be ordered $-1 \leq \lambda_1 \leq \lambda_2 \leq \ldots < \lambda_N = 1$.

- *The **algebraic connectivity** of the graph is $\mu_2 = 1 - \lambda_{N-1}$.*
- *The **generalized spectral gap** of $\boldsymbol{A}_{\text{rw}}$ is $1 - \max_{k \neq 1} |\lambda_k|$.*

The algebraic connectivity μ_2 provides a quantitative measure of how robustly connected a graph is. A larger μ_2 implies:

- The graph is "harder to disconnect" (more robust).
- Random walks on the graph mix faster, meaning they reach a stationary distribution more quickly.
- Consensus algorithms converge more rapidly.

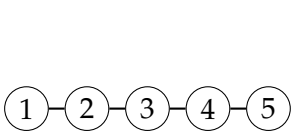
Conversely, a small μ_2 indicates that the graph is "loosely connected" or has a bottleneck, making it easier to partition.

Remark 2.2.1 (Relation with the normalized adjacency matrix)**.** *Those observations can be translated to the metrics on the normalized adjacency matrix $\boldsymbol{A}_{\text{rw}}$ of a connected graph. Indeed, since $\mu_2 = 1 - \lambda_{N-1}$, we have a more connected graph for λ_{N-1} close to 0 and a less connected graph for λ_{N-1} close to 1. The generalized spectral gap is $1 - \max_{k \neq 1} |\lambda_k|$ and corresponds to the spectral gap only if $\lambda_{N-1} > |\lambda_1|$.*

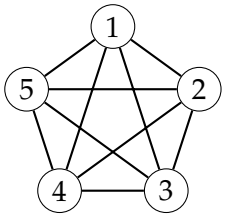
The spectral gap is a critical parameter in analyzing the rate of convergence for various network processes, including consensus algorithms and diffusion dynamics. A larger spectral gap indicates faster convergence to a stable state across the network.

2.2.4 Graph Partitions

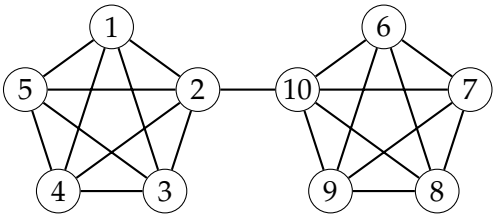
Partitions provide a way to divide a network into smaller, distinct groups of nodes, which is useful for analyzing community structures or modular behaviors within complex systems.



(a) Path Graph (P_5): $\mu_2 \approx 0.293$



(b) Complete Graph (K_5): $\mu_2 = 1.25$



(c) Two connected complete graphs connected by a single edge: $\mu_2 \approx 0.073$

Figure 2.4: Examples illustrating varying algebraic connectivity. The P_5 graph is weakly connected, resulting in a small λ_2. The K_5 graph is maximally connected, leading to a large λ_2. The two connected complete graphs are connected by a single edge leading to the presence of a bottleneck and thus a really small λ_2.

Definition 2.2.19 (Partition of a graph)**.** *A **partition** $\pi = \{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$ of the vertex set $\mathcal{V}$ is a collection of non-empty, disjoint subsets (or **cells**) such that their union covers all vertices ($\bigcup_{k=1}^{M} \mathcal{C}_k = \mathcal{V}$) and they are mutually exclusive ($\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$ for $i \neq j$).*

Definition 2.2.20 (Characteristic matrix)**.** *The **characteristic matrix** $\boldsymbol{P} \in \mathbb{R}^{N \times M}$ for a partition π is a matrix whose entries are defined by:*

$$P_{ij} = \begin{cases} 1, & \text{if vertex } i \in \mathcal{C}_j, \\ 0, & \text{otherwise.} \end{cases}$$

This matrix maps individual nodes to their respective cells. Its columns are indicator vectors for each cell, and its rows sum to 1. We will note its Moore-Penrose inverse as $\boldsymbol{P}^{\#} = (\boldsymbol{P}^\top \boldsymbol{P})^{-1} \boldsymbol{P}^\top$.

The characteristic matrix $\boldsymbol{P}$ helps to map between the full state space and the states within the partitioned substructures.

Definition 2.2.21 (Subgraphs and adjacency internal and external matrices)**.** *Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a partition $\pi = \{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$ of $\mathcal{V}$:*

- *The **induced subgraph** $\mathcal{G}_k = (\mathcal{C}_k, \mathcal{E}_k)$ for each cell $\mathcal{C}_k$ consists of only the vertices in $\mathcal{C}_k$ and all edges from $\mathcal{E}$ that connect two vertices within $\mathcal{C}_k$. The set of these* internal edges *for cell $\mathcal{C}_k$ is $\mathcal{E}_k = \{\{i, j\} \in \mathcal{E} \mid i, j \in \mathcal{C}_k\}$. The adjacency matrix of $\mathcal{G}_k$ is denoted $\boldsymbol{A}_k$.*
- ***Internal edges** are edges whose both endpoints belong to the same cell of the partition, i.e., $\mathcal{E}_\pi^I = \bigcup_{k=1}^{M} \mathcal{E}_k$. The **internal adjacency matrix** $\boldsymbol{A}_\pi^I$ is defined as $[\boldsymbol{A}_\pi^I]_{ij} = 1$ if $(i, j) \in \mathcal{E}_\pi^I$ and 0 otherwise. It is the block diagonal matrix $\boldsymbol{A}_\pi^I = \mathrm{Blockdiag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_M)$.*
- ***External edges** are edges connecting vertices that belong to different cells of the partition, i.e.,*

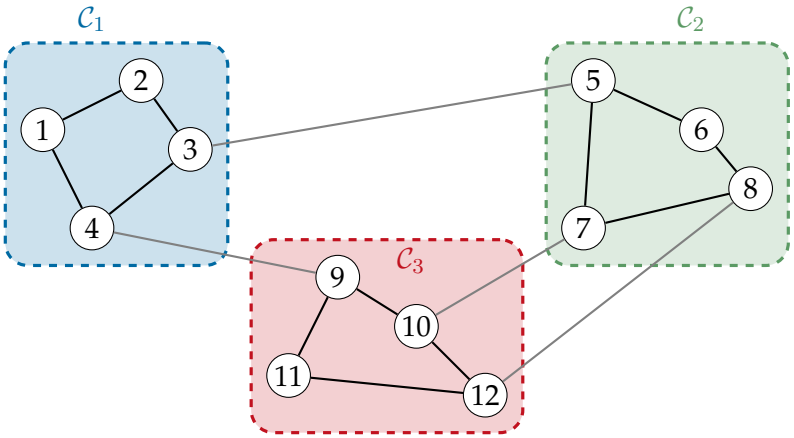


Figure 2.5: Illustration of a graph partition. A graph $\mathcal{G}$ is partitioned into three disjoint cells $\mathcal{C}_1 = \{1, 2, 3, 4\}$, $\mathcal{C}_2 = \{5, 6, 7, 8\}$, and $\mathcal{C}_3 = \{9, 10, 11, 12\}$. Edges within cells are *internal edges* in black, while those connecting different cells are *external edges* in gray.

$\mathcal{E}_\pi^E = \mathcal{E} \setminus \mathcal{E}_\pi^I$. *The* ***external adjacency matrix*** $\boldsymbol{A}_\pi^E$ *is defined as* $[\boldsymbol{A}_\pi^E]_{ij} = 1$ *if* $(i, j) \in \mathcal{E}_\pi^E$ *and 0 otherwise. We have the relation* $\boldsymbol{A} = \boldsymbol{A}_\pi^I + \boldsymbol{A}_\pi^E$.

Definition 2.2.22 (Connected partition)**.** *A partition* π *is said to be* ***connected*** *if the induced subgraphs* $\mathcal{G}_k$ *are connected for any cell* $\mathcal{C}_k \in \pi$.

Definition 2.2.23 (Selection matrix)**.** *For a partition* $\pi = \{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$ *of a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, the* ***selection matrix*** $\boldsymbol{S} \in \mathbb{R}^{N \times M}$ *is a matrix that extracts the states of agents belonging to each cell* $\mathcal{C}_k$ *from a global state vector* $\boldsymbol{x} \in \mathbb{R}^N$ *associated with the graph* $\mathcal{G}$*. Specifically, if* $\mathcal{C}_k = \{v_1, \ldots, v_{|\mathcal{C}_k|}\}$*, then* $\boldsymbol{S}_k$ *is a matrix with* 1 *at* (j, k) *and* 0 *elsewhere, such that* $\boldsymbol{S}\boldsymbol{x} = (x_{v_1}, \ldots, x_{v_{|\mathcal{C}_k|}})^\top$.

Those concepts relate to the notion of community in graph. The following example illustrates how to partition a graph and how to extract the states of agents belonging to each cell.

Example 2.2.5 (Partitioning a graph)**.** *Consider the graph from Figure 2.5 with* $N = 12$ *nodes and a partition* $\pi = \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3\}$*, where* $\mathcal{C}_1 = \{1, 2, 3, 4\}$*,* $\mathcal{C}_2 = \{5, 6, 7, 8\}$*, and* $\mathcal{C}_3 = \{9, 10, 11, 12\}$.

– *The corresponding adjacency and* ***characteristic*** *matrix are given by:*

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

– ***Induced Subgraphs and Internal Edges:*** *The induced subgraph* $\mathcal{G}_1$ *for cell* $\mathcal{C}_1 = \{1, 2, 3, 4\}$

includes nodes $\{1,2,3,4\}$ *and edges* $\{\{1,2\},\{1,4\},\{2,3\},\{3,4\}\}$. *These are the black edges fully contained within the blue region in Figure 2.5. The adjacency matrix* $\boldsymbol{A}_1$ *for* $\mathcal{G}_1$ *would be a* 4×4 *matrix.*

- ***External Edges:*** *The gray edges in Figure 2.5 are external edges. Examples include* $\{3,5\}$*, which connects* $\mathcal{C}_1$ *and* $\mathcal{C}_2$*, and* $\{4,9\}$*, connecting* $\mathcal{C}_1$ *and* $\mathcal{C}_3$.
- ***Selection Matrix:*** *For cell* $\mathcal{C}_1 = \{1,2,3,4\}$*, the selection matrix* $\boldsymbol{S}_1 \in \mathbb{R}^{4\times 12}$ *is:*

$$\boldsymbol{S}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

If $\boldsymbol{x} = (x_1, \ldots, x_{12})^\top$ *is the global state vector, then* $\boldsymbol{S}_1\boldsymbol{x} = (x_1, x_2, x_3, x_4)^\top$*, effectively extracting the states of nodes within* $\mathcal{C}_1$.

Up to a reindexing of the nodes, for a given partition π of a graph $\mathcal{G}$, the adjacency matrix $\boldsymbol{A}$ can be decomposed such that $\boldsymbol{A} = \boldsymbol{A}_\pi^I + \boldsymbol{A}_\pi^E$. Example 2.2.5 and Figure 2.5 illustrate these concepts for a sample graph.

Equitable Partitions

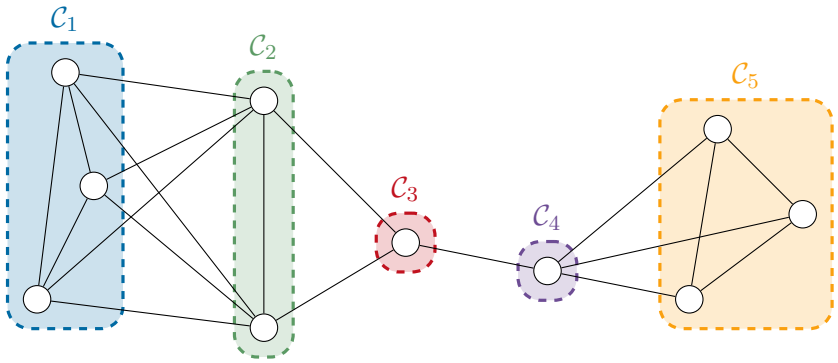
Equitable partitions are a special class of graph partitions that possess structural symmetries, simplifying the analysis of network dynamics. The definitions and results presented here are inspired by [113, Section 9.3].

Definition 2.2.24 (Equitable partition). *A partition* $\pi = \{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$ *of* $\mathcal{V}$ *is called* ***equitable*** *if for any two cells* $\mathcal{C}_i$ *and* $\mathcal{C}_j$ *(which can be the same cell), the number of neighbors in* $\mathcal{C}_j$ *of a vertex* $l \in \mathcal{C}_i$ *is a constant* b_{ij} *that is independent of the specific choice of* l *in* $\mathcal{C}_i$*. These values* b_{ij} *are called the* ***quotient graph weights****.*

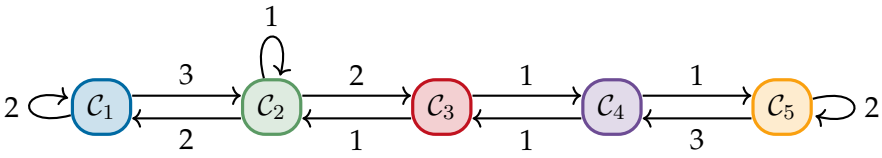
An equitable partition groups vertices that are "structurally equivalent" with respect to the partition. This symmetry allows for a reduced representation of the graph's dynamics.

Definition 2.2.25 (Quotient graph). *Given a graph* $\mathcal{G}$ *and an equitable partition* $\pi = \{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$*, the* ***quotient graph*** $\mathcal{G}/\pi$ *is a directed graph whose vertices are the cells* $\mathcal{C}_1, \ldots, \mathcal{C}_M$*. A directed edge exists from* $\mathcal{C}_i$ *to* $\mathcal{C}_j$ *with weight* b_{ij} *(as defined above) if* $b_{ij} > 0$*. The adjacency matrix of the quotient graph is* $\boldsymbol{B} = (b_{ij}) \in \mathbb{R}^{M\times M}$.

The quotient graph $\mathcal{G}/\pi$ provides a reduced-order representation of the original graph, preserving the essential connectivity patterns between the cells of an equitable partition. This reduction is particularly useful for simplifying the analysis of dynamical systems on graphs.



(a) A graph with a connected equitable partition.



(b) Quotient graph of the graph (a) with the partition $\pi = \{\mathcal{C}_1, \dots, \mathcal{C}_5\}$.

Figure 2.6: Example of a graph with equitable partition and its quotient graph.

Lemma 2.2.26 (Equitable Partition Properties). *Let $\pi = \{\mathcal{C}_1, \dots, \mathcal{C}_M\}$ be an equitable partition of a undirected graph $\mathcal{G}$ with adjacency matrix $\boldsymbol{A}$ and degree matrix $\boldsymbol{D}$. Let $\boldsymbol{P}$ be its characteristic matrix and $\boldsymbol{B}$ be the adjacency matrix of the quotient graph $\mathcal{G}/\pi$. Let $\tilde{\boldsymbol{D}} = \operatorname{diag}(d_{\mathcal{C}_1}, \dots, d_{\mathcal{C}_M})$ be the diagonal matrix of cell degrees. The following identities hold:*

(i) The adjacency matrix of the original graph $\boldsymbol{A}$ and the quotient graph $\boldsymbol{B}$ are related by $\boldsymbol{AP} = \boldsymbol{PB}$

(ii) The degree matrices of the original graph $\boldsymbol{D}$ and the quotient graph $\tilde{\boldsymbol{D}}$ are related by $\boldsymbol{DP} = \boldsymbol{P}\tilde{\boldsymbol{D}}$

(iii) The inverse of the degree matrix of the original graph $\boldsymbol{D}^{-1}$ and the quotient graph $\tilde{\boldsymbol{D}}^{-1}$ are related by $\boldsymbol{D}^{-1}\boldsymbol{P} = \boldsymbol{P}\tilde{\boldsymbol{D}}^{-1}$

Proof. (i) See [113, Lemma 9.3.1].

(ii) The (i,k)-th entry of $\boldsymbol{DP}$ is $d_i P_{ik}$. If $i \notin \mathcal{C}_k$, $P_{ik} = 0$. If $i \in \mathcal{C}_k$, the entry is $d_i = d_{\mathcal{C}_k}$. The (i,k)-th entry of $\boldsymbol{P}\tilde{\boldsymbol{D}}$ is $\sum_l P_{il}\tilde{D}_{lk} = P_{ik}\tilde{D}_{kk} = P_{ik}d_{\mathcal{C}_k}$. This is also $d_{\mathcal{C}_k}$ if $i \in \mathcal{C}_k$ and 0 otherwise.

(iii) Follows directly from (ii) by multiplying by inverses. □

2.3 Matrix Eigenvalue Theorems

This section reviews standard results regarding the eigenvalues of Hermitian matrices and thus on symmetric matrices. Specifically, we focus on variational characterizations, which allow us to bound eigenvalues using optimization over subspaces. These tools are essential for the stability and convergence analysis performed in later chapters. The following theorems are taken from [115, Chapter 4].

The first result is the Rayleigh-Ritz theorem, which relates the smallest and largest eigenvalues of a matrix to the minimum and maximum values of the Rayleigh quotient.

Rayleigh-Ritz Theorem. *[115, Theorem 4.2.2] Let $\boldsymbol{A} \in \mathbb{C}^{N\times N}$ be a Hermitian matrix with algebraically ordered eigenvalues $\lambda_1 \leq \cdots \leq \lambda_N$. The extremal eigenvalues can be expressed using the Rayleigh quotient:*

$$\lambda_1 = \lambda_{\min} = \min_{\substack{\boldsymbol{x}\in\mathbb{C}^N \\ \boldsymbol{x}\neq\boldsymbol{0}}} \frac{\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}$$

and

$$\lambda_N = \lambda_{\max} = \max_{\substack{\boldsymbol{x}\in\mathbb{C}^N \\ \boldsymbol{x}\neq\boldsymbol{0}}} \frac{\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}$$

For any non-zero vector $\boldsymbol{x} \in \mathbb{C}^N$, the Rayleigh quotient is bounded by the extremal eigenvalues:

$$\lambda_{\min} \leq \frac{\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}} \leq \lambda_{\max}$$

Equality holds if and only if $\boldsymbol{x}$ is an eigenvector corresponding to the respective eigenvalue.

[115] *Matrix Analysis*. Roger A. Horn et al. (2017).

While Rayleigh-Ritz Theorem handles the extremal cases, the Courant-Fischer theorem generalizes this concept to all eigenvalues. It characterizes the k-th eigenvalue as a min-max problem over subspaces. This is particularly useful for analyzing how structural changes in a network (such as removing edges or nodes) affect the spectral gap and, consequently, the convergence speed of consensus algorithms.

Courant-Fischer Min-Max Theorem. *[115, Theorem 4.2.6] Let $\boldsymbol{A} \in \mathbb{C}^{N\times N}$ be a Hermitian matrix with algebraically ordered eigenvalues $\lambda_1 \leq \cdots \leq \lambda_N$. For each $k \in \{1, \ldots, N\}$, the k-th eigenvalue has the following variational characterizations:*

$$\lambda_k = \min_{\substack{S\subseteq\mathbb{C}^N \\ \dim S = k}} \max_{\substack{\boldsymbol{x}\in S \\ \boldsymbol{x}\neq\boldsymbol{0}}} \frac{\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}$$

and

$$\lambda_k = \max_{\substack{S\subseteq\mathbb{C}^N \\ \dim S = N-k+1}} \min_{\substack{\boldsymbol{x}\in S \\ \boldsymbol{x}\neq\boldsymbol{0}}} \frac{\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}$$

where the optimization is over all subspaces $S \subseteq \mathbb{C}^N$ of the specified dimension.

[115] *Matrix Analysis*. Roger A. Horn et al. (2017).

Finally, we state Ostrowski's inequality. This theorem describes how the eigenvalues change under a congruence transformation $\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}^\top$. In the context of this thesis, this result helps quantify the effect of coordinate transformations into the normalized adjacency matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$ on the system dynamics.

Ostrowski's inequality. *[115, Theorem 4.5.9] Let $\boldsymbol{A}, \boldsymbol{S} \in \mathbb{R}^{N\times N}$ with $\boldsymbol{A}$ symmetric. Let the eigenvalues of $\boldsymbol{A}$, $\boldsymbol{S}\boldsymbol{S}^\top$, and $\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}^\top$ be arranged in nondecreasing order, denoted by $\lambda_k(\boldsymbol{A})$, $\lambda_k(\boldsymbol{S}\boldsymbol{S}^\top)$, and $\lambda_k(\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}^\top)$ respectively, for $k = 1, \ldots, N$. Then for each $k = 1, \ldots, N$, there exists*

a real number $\theta_k \in [\lambda_1(\boldsymbol{S}\boldsymbol{S}^\top), \lambda_N(\boldsymbol{S}\boldsymbol{S}^\top)]$ *such that* $\lambda_k(\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}^\top) = \theta_k \lambda_k(\boldsymbol{A})$.

[115] *Matrix Analysis*. Roger A. Horn et al. (2017).

2.4 Dynamical systems

Definition 2.4.1 (Autonomous Dynamical System)**.** *A* ***continuous-time autonomous dynamical system*** *is described by a system of first-order Ordinary Differential Equations (ODEs) of the form:*

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)), \tag{2.5}$$

where:

- $\mathcal{X} \subseteq \mathbb{R}^N$ *is the* ***state space*** *(the set of all possible states),*
- $\boldsymbol{x}(t) \in \mathcal{X}$ *is the* ***state vector*** *at time* t*,*
- $\boldsymbol{f} : \mathcal{X} \to \mathcal{X}$ *is a function called the* ***vector field*** *of the dynamical system.*
- *the "*$\dot{\boldsymbol{x}}(t)$*" notation denotes the time derivative of the state vector, i.e.,* $\dot{\boldsymbol{x}}(t) = \frac{d}{dt}\boldsymbol{x}(t)$*,*

The term ***autonomous*** *signifies that the function* $\boldsymbol{f}$ *does not explicitly depend on time.*

Remark 2.4.1. *For notational brevity, we will often omit the explicit time dependence of the state vector, writing* $\boldsymbol{x}$ *instead of* $\boldsymbol{x}(t)$ *when the context is clear. This convention simplifies the presentation without sacrificing formal accuracy.*

2.4.1 Existence and properties of solutions of differential equations

The definition of a dynamical system presents an abstract rule for evolution. For this abstraction to serve as a meaningful and predictive model, we must first establish several core properties of its solutions.

The most basic requirements are the **existence** and **uniqueness** of trajectories. Existence ensures that the system has a well-defined evolution from any starting point. Uniqueness is even more critical: it guarantees that the system is deterministic. A given initial state leads to one, and only one, future. This property, established under the Lipschitz condition, underpins the very concept of the system's flow and is the foundation for any predictive analysis.

Beyond determinism, a useful model must also be **robust**. In any real-world application, initial conditions are known only with finite precision, and model parameters may contain small uncertainties. It is therefore essential that solutions depend continuously on these values. A small change in the starting point or a slight variation in a system parameter should result in only a small change in the resulting trajectory over any finite time horizon.

The following definitions and propositions formalize these necessary requirements.

Definition 2.4.2 (Solution and Trajectory [116, Definition 1.3])**.** *A function* $\boldsymbol{x} : \mathbb{R}_+ \times \mathcal{X} \to \mathcal{X}$ *is a* ***solution*** *of the system (2.5) originating from an initial state* $\boldsymbol{x}_0 \in \mathcal{X}$ *if there exists a neighborhood* I *of* 0 *in* $\mathbb{R}_+$ *such that:*

- *the function* $t \in I \mapsto \boldsymbol{x}(t; \boldsymbol{x}_0) \in \mathcal{X}$ *is continuous,*
- *we have*

$$\boldsymbol{x}(t; \boldsymbol{x}_0) = \boldsymbol{x}_0 + \int_0^t \boldsymbol{f}(\boldsymbol{x}(\tau; \boldsymbol{x}_0))d\tau, \quad \forall t \in I.$$

*The point $\boldsymbol{x}_0$ is called the **initial condition** of the solution.*

*The image of this function, i.e., the set of points $\{\boldsymbol{x}(t;\boldsymbol{x}_0) \mid t \in I\}$, is called the **trajectory** or **orbit** through $\boldsymbol{x}_0$.*

[116] *Fonctions de Lyapunov : Stabilité.* Delphine Bresch-Pietri et al. (2022).

Under the assumption that the vector field function $\boldsymbol{f}$ is continuous, the solutions follow the **Carathéodory's conditions** [116]. The following proposition states the existence and uniqueness of solutions under the Lipschitz condition.

Proposition 2.4.3 (Existence and uniqueness of solutions [116, Chapter 1, Theorems 1, 2 and 4]). *Let $\boldsymbol{f} : \mathcal{X} \to \mathcal{X}$ be a continuous function. Then, for any $\boldsymbol{x}_0 \in \mathcal{X}$, there exists a neighborhood I of 0 in $\mathbb{R}_+$ and a function $t \mapsto \boldsymbol{x}(t;\boldsymbol{x}_0) \in \mathcal{X}$ that is a solution of the system (2.5) with initial condition $\boldsymbol{x}_0$ on the interval I. This solution is C^1 on I and verify*

$$\dot{\boldsymbol{x}}(t;\boldsymbol{x}_0) = \boldsymbol{f}(\boldsymbol{x}(t;\boldsymbol{x}_0)), \quad \forall t \in I.$$

*Moreover, if the function $\boldsymbol{f}$ satisfies the **Lipschitz condition** over $\mathcal{X}$, i.e.,*

$$\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{y})\| \leq K \|\boldsymbol{x} - \boldsymbol{y}\|, \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X},$$

for some $K > 0$, then the solution is unique and can be extended to the entire positive real line. i.e., for any $\boldsymbol{x} \in \mathcal{X}$, there exists a unique solution $\boldsymbol{x}(t;\boldsymbol{x}_0)$ for all $t \in \mathbb{R}_+$ passing through $\boldsymbol{x}$.

[116] *Fonctions de Lyapunov : Stabilité.* Delphine Bresch-Pietri et al. (2022).

Proposition 2.4.4 (Continuity in parameters of solutions [117, Theorem 3.5]). *Let $\boldsymbol{f} : \mathcal{X} \times \mathbb{R} \to \mathcal{X}$ be a continuous function and locally Lipschitz in its first argument uniformly in the second. For any $(\boldsymbol{x}_0, \alpha) \in \mathcal{X} \times \mathbb{R}$, let $\boldsymbol{x}(t;\boldsymbol{x}_0,\alpha)$ be the solution of the system (2.5) with initial condition $\boldsymbol{x}_0$ on an interval $I \subseteq \mathbb{R}_+$. Then, the solution $\boldsymbol{x}(t;\boldsymbol{x}_0,\alpha)$ is continuous in the parameters $(\boldsymbol{x}_0, \alpha)$ on the interval $I \times \mathbb{R}$. i.e., for all $\varepsilon > 0$, there exists a $\delta > 0$ such that if*

$$\left\|\boldsymbol{x}_0 - \boldsymbol{x}_0'\right\| < \delta \quad \text{and} \quad \left|\alpha - \alpha'\right| < \delta,$$

then there exists a unique solution $\boldsymbol{x}'(t;\boldsymbol{x}_0',\alpha')$ of the system (2.5) with initial condition $\boldsymbol{x}_0'$ on the interval I and it verifies

$$\left\|\boldsymbol{x}'(t;\boldsymbol{x}_0',\alpha') - \boldsymbol{x}(t;\boldsymbol{x}_0,\alpha)\right\| < \varepsilon, \quad \forall t \in I.$$

[117] *Nonlinear Systems.* Hassan K Khalil. (2002).

The collection of all possible solutions can be unified into a single mathematical object called the flow. Assuming conditions that guarantee unique solutions for all time (cf. Proposition 2.4.3), the flow describes the mapping from any initial state to its future state.

Definition 2.4.5 (Flow of a Dynamical System [118]). *Let $\boldsymbol{x}(t;\boldsymbol{x}_0)$ denote the unique solution to (2.5) with initial condition $\boldsymbol{x}(0) = \boldsymbol{x}_0$. The **flow** (sometimes called the **semiflow**) of the dynamical system is the map $\phi : \mathbb{R}_+ \times \mathcal{X} \to \mathcal{X}$ defined by*

$$\phi(t, \boldsymbol{x}_0) = \boldsymbol{x}(t;\boldsymbol{x}_0).$$

It is often convenient to use the shorthand notation $\phi_t(\boldsymbol{x}_0) = \phi(t, \boldsymbol{x}_0)$. The flow has the following properties:

(i) ***Identity Property:*** *$\phi_0(\boldsymbol{x}) = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{X}$. i.e., the flow at time zero is the identity map.*

(ii) ***Semigroup Property:*** *$\phi_{t+s}(\boldsymbol{x}) = \phi_t(\phi_s(\boldsymbol{x}))$ for all $\boldsymbol{x} \in \mathcal{X}$ and for all $t, s \geq 0$.*

[118] "The Dynamical Systems Approach to Differential Equations". Morris W. Hirsch. (1984).

2.4.2 Invariant Sets and Asymptotic Behavior

A central goal in the study of dynamical systems is to characterize how trajectories behave over time. This begins with the core concept of an invariant set: a region of the state space that traps trajectories within it for all time.

Definition 2.4.6 (Positively Invariant Set [116, Definition 1.65])**.** *Consider the autonomous system (2.5) defined on a domain $\mathcal{X} \subseteq \mathbb{R}^N$. A set $\mathcal{D} \subseteq \mathcal{X}$ is said to be* ***positively invariant*** *(or* ***forward invariant****) with respect to the system if every trajectory starting in $\mathcal{D}$ remains in $\mathcal{D}$ for all future time. Formally,*

$$\text{if } \boldsymbol{x}_0 \in \mathcal{D}, \quad \text{then } \phi_t(\boldsymbol{x}_0) \in \mathcal{D} \text{ for all } t \geq 0.$$

[116] *Fonctions de Lyapunov : Stabilité.* Delphine Bresch-Pietri et al. (2022).

While the definition of positive invariance is conceptually clear, checking it for every possible initial condition is intractable. A more powerful approach is to derive a local condition on the vector field $\boldsymbol{f}$ that guarantees this global property. The key insight is that a trajectory can only exit a set $\mathcal{S}$ by crossing its boundary, $\partial\mathcal{S}$.

Therefore, for $\mathcal{S}$ to be positively invariant, it is sufficient to ensure that at every point $\boldsymbol{x}$ on the boundary, the direction of motion, given by the vector field $\boldsymbol{f}(\boldsymbol{x})$, does not point outward. The vector $\boldsymbol{f}(\boldsymbol{x})$ must either point into the interior of $\mathcal{S}$ or be tangent to its boundary. To formalize this geometric condition for general sets, including those with non-smooth boundaries (e.g., corners), we introduce the tangent cone and Nagumo's theorem.

Definition 2.4.7 (Tangent cone [119])**.** *Let $\mathcal{S} \subseteq \mathcal{X}$ be a compact set. The tangent cone of $\mathcal{S}$ at $\boldsymbol{x} \in \mathcal{S}$ is the set*

$$\mathcal{T}_{\mathcal{S}}(\boldsymbol{x}) = \left\{ \boldsymbol{y} \in \mathcal{X} \mid \liminf_{h \to 0} \frac{d(\boldsymbol{x} + h\boldsymbol{y}, \mathcal{S})}{h} = 0 \right\}.$$

Nagumo's Theorem ([119])**.** *Consider the system of differential equations $\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t))$, and assume that, for each initial condition $\boldsymbol{x}_0 \in \mathcal{X}$, it admits a globally unique solution. Let $\mathcal{S} \subseteq \mathcal{X}$ be a closed and convex set. Then, the set $\mathcal{S}$ is forward invariant for the system if and only If*

$$f(\boldsymbol{x}) \in \mathcal{T}_{\mathcal{S}}(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \in \mathcal{S}.$$

[119] "Set Invariance in Control". F. Blanchini. (1999).

Having defined invariance, we can now introduce the elementary structures that describe long-term behavior, which are themselves prime examples of invariant sets.

Definition 2.4.8 (Equilibrium Point)**.** *A point $\boldsymbol{x}^* \in \mathcal{X}$ is an* ***equilibrium point*** *(also called a* ***fixed point*** *or* ***stationary point****) of the system (2.5) if the vector field vanishes at that point: $\boldsymbol{f}(\boldsymbol{x}^*) = \boldsymbol{0}$. An equilibrium point corresponds to a constant solution, $\boldsymbol{x}(t) = \boldsymbol{x}^*$ for all t.*

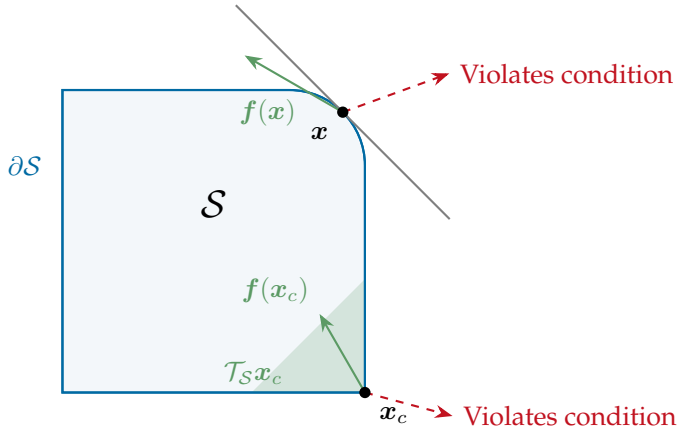


Figure 2.7: Illustration of Nagumo's condition for a set $\mathcal{S} \subseteq \mathcal{X}$ with a smooth boundary and a sharp corner. For invariance, the vector field $f(\boldsymbol{x})$ must belong to the tangent cone $\mathcal{T}_S\boldsymbol{x}$ at every boundary point $\boldsymbol{x} \in \partial\mathcal{S}$. The red dashed vector are examples of vectors that violates the condition.i.e., a dynamical system with a vector field that is not in the tangent cone at a boundary point will not make the set forward invariant.

Definition 2.4.9 (Periodic Orbit)**.** *A trajectory $\boldsymbol{x}(t)$ is a **periodic orbit** if there exists a minimal time $T > 0$, called the **period**, such that:*

- $\boldsymbol{x}(T) = \boldsymbol{x}(0)$,
- *The initial state $\boldsymbol{x}(0)$ is not an equilibrium point.*

*The set of points $\mathcal{O} = \{\boldsymbol{x}(t) \mid 0 \leq t < T\}$ forms a closed curve in the state space and is a positively invariant set. A periodic orbit that is isolated from other periodic orbits is called a **limit cycle**.*

More generally, the long-term destination of any single trajectory is captured by its omega-limit set, which is guaranteed to be an invariant set.

Definition 2.4.10 (ω-limit set)**.** *The **omega-limit set** of a point $\boldsymbol{x}_0 \in \mathcal{X}$, denoted $\omega(\boldsymbol{x}_0)$, is the set of all points $\boldsymbol{y} \in \mathcal{X}$ for which there exists a sequence of times $t_k \to \infty$ such that the trajectory $\phi_{t_k}(\boldsymbol{x}_0)$ converges to $\boldsymbol{y}$.*

$$\omega(\boldsymbol{x}_0) = \left\{ \boldsymbol{y} \in \mathcal{X} \mid \exists \{t_k\}_{k\in\mathbb{N}} \text{ with } t_k \to \infty \text{ s.t. } \lim_{k\to\infty} \phi_{t_k}(\boldsymbol{x}_0) = \boldsymbol{y} \right\}.$$

The ω-limit set of any point is a closed and positively invariant set.

While the ω-limit set describes the asymptotic destination of a single trajectory, the concept of an attracting set describes the collective convergence of a whole neighborhood of trajectories to a common invariant set. This leads to the core concepts of attracting sets and their basins of attraction.

Definition 2.4.11 (Attracting Set)**.** *Let $\mathcal{A} \subset \mathcal{X}$ be a compact set. $\mathcal{A}$ is called an **attracting set** for the system (2.5) if it is positively invariant and there exists an open neighborhood $\mathcal{U}$ containing $\mathcal{A}$ such that for every initial condition $\boldsymbol{x}_0 \in \mathcal{U}$, the resulting trajectory converges to $\mathcal{A}$. Formally,*

$$\forall \boldsymbol{x}_0 \in \mathcal{U}, \quad \lim_{t\to\infty} d(\phi_t(\boldsymbol{x}_0), \mathcal{A}) = 0,$$

where $d(\boldsymbol{y}, \mathcal{A}) = \inf_{\boldsymbol{z}\in\mathcal{A}} \|\boldsymbol{y} - \boldsymbol{z}\|$ is the point-to-set distance.

Definition 2.4.12 (Basin of Attraction)**.** *Let $\mathcal{A}$ be an attracting set. The **basin of attraction** of $\mathcal{A}$, denoted $\mathcal{B}(\mathcal{A})$, is the set of all initial conditions in $\mathcal{X}$ whose trajectories converge to $\mathcal{A}$.*

$$\mathcal{B}(\mathcal{A}) = \left\{ \boldsymbol{x}_0 \in \mathcal{X} \mid \lim_{t\to\infty} d(\phi_t(\boldsymbol{x}_0), \mathcal{A}) = 0 \right\}.$$

Equivalently, the basin of attraction is the set of points whose omega-limit set is contained within $\mathcal{A}$:

$$\mathcal{B}(\mathcal{A}) = \{\boldsymbol{x}_0 \in \mathcal{X} \mid \omega(\boldsymbol{x}_0) \subseteq \mathcal{A}\}.$$

Remark 2.4.2. *The basin of attraction $\mathcal{B}(\mathcal{A})$ is an open and positively invariant set. Its boundary, $\partial\mathcal{B}(\mathcal{A})$, is known as the **separatrix**. A particularly important case is when the attracting set is an asymptotically stable equilibrium point $\boldsymbol{x}^*$; then $\mathcal{B}(\boldsymbol{x}^*)$ defines the set of initial conditions from which the system will return to equilibrium. If $\mathcal{B}(\mathcal{A}) = \mathcal{X}$, then $\mathcal{A}$ is said to be **globally attracting**.*

2.4.3 Cooperative, monotone, and strongly monotone systems

Beyond the general properties of invariance and asymptotic behavior, certain classes of systems exhibit a much higher degree of structural regularity in their flows. A particularly important class are **cooperative systems**, whose flows are order-preserving. This geometric property imposes powerful constraints on the dynamics, precluding complex oscillatory or chaotic behavior. As we will see, this structured behavior has profound consequences for the system's long-term evolution, often simplifying it to convergence towards equilibria and making its analysis significantly more tractable than that of general nonlinear systems.

Definition 2.4.13. *We use the following notation for the standard partial order on $\mathbb{R}^N$:*

- *$\boldsymbol{x} \leq \boldsymbol{y}$ means $x_i \leq y_i$ for all components $i = 1, \ldots, N$.*
- *$\boldsymbol{x} < \boldsymbol{y}$ means $\boldsymbol{x} \leq \boldsymbol{y}$ and $\boldsymbol{x} \neq \boldsymbol{y}$.*
- *$\boldsymbol{x} \ll \boldsymbol{y}$ means $x_i < y_i$ for all components $i = 1, \ldots, N$.*

Cooperative systems

The starting point is the definition of a cooperative system, which generates a monotone flow.

Definition 2.4.14 (Cooperative System)**.** *A dynamical system (2.5) is **cooperative** if its flow ϕ is order-preserving with respect to the standard partial order $\leq$ on $\mathbb{R}^N$. That is, for any pair of initial conditions $\boldsymbol{x}_0, \boldsymbol{y}_0 \in \mathcal{X}$:*

$$\boldsymbol{x}_0 \leq \boldsymbol{y}_0 \quad \Longrightarrow \quad \phi_t(\boldsymbol{x}_0) \leq \phi_t(\boldsymbol{y}_0) \quad \text{for all } t \geq 0.$$

*In this case, we say that the flow ϕ is **monotone**.*

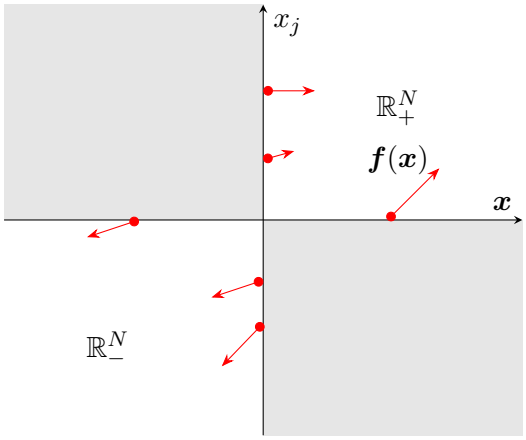


Figure 2.8: A cooperative system with a monotone flow. The red arrows represent the vector field of the system. The shaded regions represent the unordered orthant.

Whether a system is cooperative can often be verified directly from its vector field $\boldsymbol{f}$. The central condition is the **Kamke–Müller condition** or, equivalently, the **quasimonotone condition** of the vector field $\boldsymbol{f}$ [120].

Definition 2.4.15 (Quasimonotone function)**.** *A function $\boldsymbol{f} : \mathcal{X} \to \mathcal{X}$ is **quasimonotone increasing** if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ with $\boldsymbol{x} \leq \boldsymbol{y}$, and for any index $i \in \{1, \ldots, N\}$:*

$$\text{if } x_i = y_i, \quad \text{then } f_i(\boldsymbol{x}) \leq f_i(\boldsymbol{y}). \tag{2.6}$$

In other words, increasing other components x_j ($j \neq i$) cannot decrease the value of the i-th component, $f_i(\boldsymbol{x})$.

Proposition 2.4.16 (Kamke–Müller condition for cooperativity [120, Theorem 3.2])**.** *A dynamical system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ is cooperative if and only if the function $\boldsymbol{f} : \mathcal{X} \to \mathcal{X}$ is* quasimonotone increasing.

[120] "Chapter 4 Monotone Dynamical Systems". M.W. Hirsch et al. (2006).

Proof. The proof follows the classical approach in [120, Theorem 3.2], extended here to cover non-differentiable vector fields.

($\Leftarrow$) Assume $\boldsymbol{f}$ is quasimonotone increasing. Let $\boldsymbol{x}_0 \leq \boldsymbol{y}_0$. We must show that $\phi_t(\boldsymbol{x}_0) \leq \phi_t(\boldsymbol{y}_0)$ for all $t \geq 0$. Let $\boldsymbol{x}(t) = \phi_t(\boldsymbol{x}_0)$. Consider a perturbed vector field $\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{f}(\boldsymbol{y}) + \boldsymbol{\epsilon}$ for a constant vector $\boldsymbol{\epsilon} \gg \boldsymbol{0}$ (i.e., $\epsilon_i > 0$ for all i). Let $\boldsymbol{y}_\epsilon(t)$ be the solution to $\dot{\boldsymbol{y}} = \boldsymbol{g}(\boldsymbol{y})$ with initial condition $\boldsymbol{y}_\epsilon(0) = \boldsymbol{y}_0$.

We claim that $\boldsymbol{x}(t) < \boldsymbol{y}_\epsilon(t)$ for all $t > 0$ in their common interval of existence. Suppose the claim is false. Let $\tau = \inf\{t > 0 \mid \boldsymbol{x}(t) \not< \boldsymbol{y}_\epsilon(t)\}$. We first establish that $\tau > 0$. At $t = 0$, we have $\boldsymbol{x}(0) = \boldsymbol{x}_0 \leq \boldsymbol{y}_0 = \boldsymbol{y}_\epsilon(0)$.

- For any component i where $x_{0,i} < y_{0,i}$, continuity of solutions ensures that $x_i(t) < y_{\epsilon,i}(t)$ for t in some interval $[0, \delta_i)$ for some $\delta_i > 0$.
- For any component i where $x_{0,i} = y_{0,i}$, we have $\boldsymbol{x}_0 \leq \boldsymbol{y}_0$ and $x_{0,i} = y_{0,i}$. By the quasimonotone condition, $f_i(\boldsymbol{x}_0) \leq f_i(\boldsymbol{y}_0)$. The time derivatives at $t = 0$ are $\dot{x}_i(0) = f_i(\boldsymbol{x}_0)$ and

[120] "Chapter 4 Monotone Dynamical Systems". M.W. Hirsch et al. (2006).

$\dot{y}_{\epsilon,i}(0) = f_i(\boldsymbol{y}_0) + \epsilon_i$. Thus, the derivative of the difference $h_i(t) = y_{\epsilon,i}(t) - x_i(t)$ at $t = 0$ is

$$\dot{h}_i(0) = \dot{y}_{\epsilon,i}(0) - \dot{x}_i(0) = (f_i(\boldsymbol{y}_0) - f_i(\boldsymbol{x}_0)) + \epsilon_i \geq \epsilon_i > 0.$$

Since $h_i(0) = 0$ and $\dot{h}_i(0) > 0$, it follows that $h_i(t) > 0$ (i.e., $x_i(t) < y_{\epsilon,i}(t)$) for t in some interval $(0, \delta_i)$.

Combining both cases, we see that $\boldsymbol{x}(t) < \boldsymbol{y}_\epsilon(t)$ holds for all $t \in (0, \delta)$ for some $\delta > 0$. This proves that $\tau \geq \delta > 0$.

By the definition of τ and the continuity of solutions, we must have $\boldsymbol{x}(\tau) \leq \boldsymbol{y}_\epsilon(\tau)$, and there must be at least one component i for which $x_i(\tau) = y_{\epsilon,i}(\tau)$. For this index i, the function $h_i(t) = y_{\epsilon,i}(t) - x_i(t)$ satisfies $h_i(t) > 0$ for $t \in (0, \tau)$ and $h_i(\tau) = 0$. Let us evaluate the derivative of $h_i(t)$ at $t = \tau$:

$$\dot{h}_i(\tau) = \dot{y}_{\epsilon,i}(\tau) - \dot{x}_i(\tau) = (f_i(\boldsymbol{y}_\epsilon(\tau)) + \epsilon_i) - f_i(\boldsymbol{x}(\tau)).$$

Since $\boldsymbol{x}(\tau) \leq \boldsymbol{y}_\epsilon(\tau)$ and $x_i(\tau) = y_{\epsilon,i}(\tau)$, the quasimonotone condition (2.6) implies $f_i(\boldsymbol{x}(\tau)) \leq f_i(\boldsymbol{y}_\epsilon(\tau))$. Therefore,

$$\dot{h}_i(\tau) = (f_i(\boldsymbol{y}_\epsilon(\tau)) - f_i(\boldsymbol{x}(\tau))) + \epsilon_i \geq \epsilon_i > 0.$$

The fact that $h_i(\tau) = 0$ and $\dot{h}_i(\tau) > 0$ implies that for some small $\delta' > 0$, we must have $h_i(t) < 0$ for all $t \in (\tau - \delta', \tau)$. i.e., $y_{\epsilon,i}(t) < x_i(t)$ for all $t \in (\tau - \delta', \tau)$. This contradicts the definition of τ, because it requires $x_i(t) < y_{\epsilon,i}(t)$ for all $t \in (0, \tau)$. Thus, our initial assumption that the claim is false must be incorrect.

So, we have shown that $\phi_t(\boldsymbol{x}_0) < \boldsymbol{y}_\epsilon(t)$ for all $t > 0$. Taking the limit as $\boldsymbol{\epsilon} \to \mathbf{0}$, the solution $\boldsymbol{y}_\epsilon(t)$ converges to $\phi_t(\boldsymbol{y}_0)$ due to the continuous dependence of solutions on parameters. By the closed nature of the partial order, we conclude that $\phi_t(\boldsymbol{x}_0) \leq \phi_t(\boldsymbol{y}_0)$.

($\Rightarrow$) Assume the flow is monotone. Let $\boldsymbol{x} \leq \boldsymbol{y}$ with $x_i = y_i$ for some i. By monotonicity, $\phi_t(\boldsymbol{x}) \leq \phi_t(\boldsymbol{y})$ for all $t \geq 0$. Define the difference $\boldsymbol{z}(t) = \phi_t(\boldsymbol{y}) - \phi_t(\boldsymbol{x})$. We have $\boldsymbol{z}(t) \geq \mathbf{0}$ for $t \geq 0$ and $z_i(0) = 0$. This implies that the right-derivative of $z_i(t)$ at $t = 0$ must be non-negative:

$$\dot{z}_i(0) = \lim_{h \to 0^+} \frac{z_i(h) - z_i(0)}{h} = \lim_{h \to 0^+} \frac{z_i(h)}{h} \geq 0.$$

But $\dot{z}_i(0) = \dot{y}_i(0) - \dot{x}_i(0) = f_i(\boldsymbol{y}) - f_i(\boldsymbol{x})$. Therefore, $f_i(\boldsymbol{y}) \geq f_i(\boldsymbol{x})$, which is the quasimonotone condition. □

When $\boldsymbol{f}$ is differentiable, this reduces to a simple check on the Jacobian.

Corollary 2.4.17 (Metzler Jacobian Criterion [121, Chapter 4, Theorem 1.1])**.** *A smooth dynamical system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ is cooperative if and only if its Jacobian matrix $\boldsymbol{J}(\boldsymbol{x}) = \partial \boldsymbol{f} / \partial \boldsymbol{x}$ is a **Metzler matrix** for all $\boldsymbol{x} \in \mathcal{X}$. A matrix is Metzler if all of its off-diagonal entries are non-negative:*

$$J_{ij}(\boldsymbol{x}) = \frac{\partial f_i}{\partial x_j}(\boldsymbol{x}) \geq 0 \quad \textit{for all } i \neq j.$$

[121] *Monotone Dynamical Systems*. Hal L. Smith. (2008).

Even this baseline property imposes meaningful constraints on the dynamics. Trajectories cannot alternate between rising and falling intervals, and periodic orbits cannot be ordered.

This raises a natural question: if cooperativity preserves order, can one strengthen the property so that inequalities between trajectories not only persist but also become strict over time? This leads to the notion of *strong monotonicity*.

Proposition 2.4.18 (Dynamical Consequences of Cooperativity [120, Proposition 1.5, 1.6])**.** *A non-trivial periodic orbit of a cooperative system is* ***unordered****. That is, for any two distinct points* $\boldsymbol{x}, \boldsymbol{y}$ *on the orbit, neither* $\boldsymbol{x} \leq \boldsymbol{y}$ *nor* $\boldsymbol{y} \leq \boldsymbol{x}$ *can hold.*

Consequently, cooperative systems do not admit limit cycles.

[120] "Chapter 4 Monotone Dynamical Systems". M.W. Hirsch et al. (2006).

In summary, cooperative systems behave in a more structured way than general nonlinear systems. Their dynamics resemble those of one-dimensional flows, with trajectories typically converging toward equilibria rather than displaying recurrent or chaotic patterns.

Strongly monotone systems

Cooperativity ensures that order is preserved, but it does not guarantee that inequalities become strict. A stronger property, called *strong monotonicity*, requires that once two initial conditions are distinct, their separation becomes strict under the flow.

Definition 2.4.19 (Strongly Monotone Flow)**.** *A flow* ϕ *is*

- ***strongly monotone*** *if for any* $\boldsymbol{x}_0, \boldsymbol{y}_0 \in \mathcal{X}$ *with* $\boldsymbol{x}_0 < \boldsymbol{y}_0$*, we have* $\phi_t(\boldsymbol{x}_0) \ll \phi_t(\boldsymbol{y}_0)$ *for all* $t > 0$*,*
- ***eventually strongly monotone*** *if for any* $\boldsymbol{x}_0, \boldsymbol{y}_0 \in \mathcal{X}$ *with* $\boldsymbol{x}_0 < \boldsymbol{y}_0$*, there exists a time* $t_0 > 0$ *such that* $\phi_{t_0}(\boldsymbol{x}_0) \ll \phi_{t_0}(\boldsymbol{y}_0)$*.*

Clearly, a flow that is strongly monotone is also eventually strongly monotone, but the converse is not true in general.

Strong monotonicity typically arises when cooperativity is combined with an *irreducibility* condition, which ensures that all variables influence one another, directly or indirectly. For systems with a continuously differentiable vector field $\boldsymbol{f}$, strong monotonicity can be verified by examining the Jacobian matrix.

Proposition 2.4.20 (Sufficient Condition for Strong Monotonicity [121, Chapter 4, Theorem 1.1])**.** *Let* $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ *be a cooperative system with a continuously differentiable vector field* $\boldsymbol{f}$ *on a convex domain* $\mathcal{X}$*. If the Jacobian matrix* $\boldsymbol{J}(\boldsymbol{x})$ *is irreducible for all* $\boldsymbol{x} \in \mathcal{X}$*, then the flow* ϕ_t *is strongly monotone.*

[121] *Monotone Dynamical Systems*. Hal L. Smith. (2008).

For non differentiable vector fields, we can introduce a novel object as the *dependency graph* of the dynamics and define a more general notion of irreducibility on this object.

Definition 2.4.21 (Dependency Graph and Irreducibility)**.** *For a dynamical system* $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ *on a domain* $\mathcal{X} \subseteq \mathbb{R}^N$*, the* ***dependency graph****, denoted* $\mathcal{G}(\boldsymbol{f})$*, is a directed graph* $(\mathcal{V}, \mathcal{E})$ *where:*

- *The set of vertices is* $\mathcal{V} = \{1, 2, \ldots, N\}$*, corresponding to the components of the state vector* $\boldsymbol{x}$*.*
- *A directed edge* (j, i) *exists from vertex* j *to vertex* i *if the function* f_i *depends on the variable* x_j*.*

For a general continuous vector field $\boldsymbol{f}$*, the condition for the existence of an edge* (j, i) *is made precise as follows: for any point* $\boldsymbol{u} \in \mathcal{X}$ *and any non-degenerate interval* $[a, b]$ *such that the line*

segment

$$L = \{(u_1, \dots, u_{j-1}, \xi, u_{j+1}, \dots, u_N) \mid \xi \in [a, b]\}$$

is contained in $\mathcal{X}$, the function $\xi \mapsto f_i(\boldsymbol{v})$ for $\boldsymbol{v} \in L$ is not constant on $[a, b]$.

*A cooperative system is **irreducible** if its dependency graph $G(\boldsymbol{f})$ is strongly connected.*

This means that no variable evolves independently: a change in any one component eventually affects all the others.

Theorem 2.4.22 (Irreducibility Implies Strong Monotonicity)**.** *Let $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ define a cooperative and irreducible dynamical system on a domain $\mathcal{X} \subseteq \mathbb{R}^N$. Then the flow ϕ_t is strongly monotone.*

Proof. Let $\boldsymbol{x}_0, \boldsymbol{y}_0 \in \mathcal{X}$ be initial conditions such that $\boldsymbol{x}_0 < \boldsymbol{y}_0$. Let $\boldsymbol{x}(t) = \phi_t(\boldsymbol{x}_0)$ and $\boldsymbol{y}(t) = \phi_t(\boldsymbol{y}_0)$, and define the difference vector $\boldsymbol{z}(t) = \boldsymbol{y}(t) - \boldsymbol{x}(t)$. By Proposition 2.4.16, a cooperative system ensures that the flow is quasimonotone, so $\boldsymbol{z}(t) \geq \boldsymbol{0}$ for all $t \geq 0$. Our goal is to prove that $\boldsymbol{z}(t) \gg \boldsymbol{0}$ for all $t > 0$.

The argument relies on showing that the set of components for which $z_i(t)$ is strictly positive expands over time, following the paths in the dependency graph $G(\boldsymbol{f})$. Let $\mathcal{I}(0) = \{i \mid z_i(0) > 0\}$. Since $\boldsymbol{x}_0 < \boldsymbol{y}_0$, this set is non-empty. We proceed by induction on the length of the shortest path from $\mathcal{I}(0)$ to any other component.

First, consider a component k such that there is a direct edge (j, k) from some $j \in \mathcal{I}(0)$. Since $k \notin \mathcal{I}(0)$, we have $z_k(0) = 0$. The evolution of $z_k(t)$ is given by the integral

$$z_k(t) = \int_0^t [f_k(\boldsymbol{y}(\tau)) - f_k(\boldsymbol{x}(\tau))]\, d\tau \coloneqq \int_0^t g_k(\tau)\, d\tau,$$

The quasimonotone condition (2.6) guarantees that $g_k(\tau) \geq 0$. To show that $z_k(t) > 0$ for $t > 0$, we must demonstrate that $g_k(\tau)$ is not identically zero on any interval $[0, t]$ for $t > 0$.

Assume for contradiction that $g_k(\tau) = 0$ for all $\tau \in [0, t]$ for some $t > 0$. This implies $z_k(\tau) = 0$ for this interval. Since $j \in \mathcal{I}(0)$, $z_j(0) > 0$, and by continuity of solutions, there exists a $t' \in (0, t]$ such that $z_j(\tau) > 0$ for all $\tau \in [0, t']$. For any fixed $\tau \in (0, t']$, we have a state where $x_j(\tau) < y_j(\tau)$ and $x_k(\tau) = y_k(\tau)$, yet $f_k(\boldsymbol{y}(\tau)) = f_k(\boldsymbol{x}(\tau))$. By (2.6), f_k must be constant along the line segment between $\boldsymbol{x}(\tau)$ and $\boldsymbol{y}(\tau)$. This contradicts the irreducibility condition for the edge (j, k), which requires that f_k not be constant on the non-degenerate interval $[x_j(\tau), y_j(\tau)]$ when all other components are held fixed. Thus, the assumption is false, and it must be that $z_k(t) > 0$ for all $t > 0$.

This argument extends inductively. Assume that for any component i reachable from $\mathcal{I}(0)$ by a path of length at most m, we have $z_i(t) > 0$ for all $t > 0$. Now, consider a component k whose shortest path from $\mathcal{I}(0)$ has length $m + 1$. It has an immediate predecessor, say j, whose shortest path has length at most m. By the inductive hypothesis, $z_j(t) > 0$ for all $t > 0$. The same argument as in the base case applies: if $z_k(t)$ were to be zero for any positive duration, it would imply that f_k is constant with respect to changes in the j-th component over a non-degenerate interval, again contradicting irreducibility.

Since the dependency graph $G(\boldsymbol{f})$ is strongly connected, every component $i \in \{1, \dots, N\}$ is reachable from the non-empty set $\mathcal{I}(0)$. By induction on the path length, it follows that $z_i(t) > 0$ for all i and for all $t > 0$. Therefore, $\boldsymbol{z}(t) \gg 0$, which is equivalent to $\phi_t(\boldsymbol{x}_0) \ll \phi_t(\boldsymbol{y}_0)$ for all $t > 0$. □

In conclusion, strongly monotone systems sharpen the picture provided by cooperativity. Ordered states remain strictly separated throughout their evolution, and the influence of each variable propagates across the entire system. As a result, their asymptotic behavior is even more constrained and predictable, often reducing to convergence toward equilibria.

We will see next that the stronger property of strong monotonicity also has consequences beyond directly ordered trajectories. This motivates the broader concept of *strongly order-preserving* flows, which extend the same rigidity to nearby states.

Strongly Order-Preserving Flows and Their Consequences

A *strongly order-preserving* (SOP) flow extends the idea of monotonicity by requiring that order, once present, also propagates robustly to nearby states.

Definition 2.4.23 (Strongly Order-Preserving Flows)**.** *A flow ϕ is **strongly order-preserving (SOP)** if it is monotone and whenever $\boldsymbol{x} < \boldsymbol{y}$ there exists open subsets U and V with $\boldsymbol{x} \in U$ and $\boldsymbol{y} \in V$ and a $t_0 > 0$ such that*

$$\phi_{t_0}(U) \leq \phi_{t_0}(V).$$

Moreover, if ϕ is monotone, then $\phi_t(U) \leq \phi_t(V)$ for all $t \geq t_0$.

By continuity, any strongly monotone flow is automatically SOP, since one can choose $t_0 > 0$ and suitable neighborhoods to satisfy the definition.

Proposition 2.4.24 (Strongly Monotone Flows are SOP [121, Chapter 2, Proposition 1.1])**.** *If ϕ is eventually strongly monotone, then it is SOP.*

[121] *Monotone Dynamical Systems*. Hal L. Smith. (2008).

The first consequence is that ω-limit sets of SOP flows cannot contain points that are ordered.

Proposition 2.4.25 (Nonordering of Omega Limit Sets [120, Theorem 1.8])**.** *Let $\omega(\boldsymbol{x})$ be an omega limit set of a monotone flow ϕ.*

(i) No points of $\omega(\boldsymbol{x})$ are related by $\ll$.

(ii) If $\omega(\boldsymbol{x})$ is a periodic orbit, or ϕ is SOP, then no points of $\omega(\boldsymbol{x})$ is related by $<$.

A second, and more profound, consequence is the *Limit Set Dichotomy*, which restricts the possible long-term relations between two ordered initial states.

Theorem 2.4.26 (Limit Set Dichotomy [120, Theorem 1.16])**.** *Let ϕ be a SOP and let all orbits have compact closure. For any two initial states $\boldsymbol{x}_0, \boldsymbol{y}_0 \in \mathcal{X}$ such that $\boldsymbol{x}_0 < \boldsymbol{y}_0$, exactly one of the following must hold:*

(i) The omega-limit sets are strictly ordered: $\omega(\boldsymbol{x}_0) < \omega(\boldsymbol{y}_0)$.

(ii) The omega-limit sets are identical and consist only of equilibria.

[120] "Chapter 4 Monotone Dynamical Systems". M.W. Hirsch et al. (2006).

Taken together, these results show that strongly order-preserving flows inherit and extend the rigidity of monotone dynamics. By excluding both oscillatory patterns and complex attractors, they channel the long-term evolution of trajectories toward simple outcomes, typically convergence to equilibria. In this way, SOP flows provide a unifying framework: they capture

the structure imposed by strong monotonicity while applying in a broader setting, making them a central tool for understanding organisation of trajectories in nonlinear dynamics.

2.4.4 Lyapunov Stability of Autonomous Systems

Stability theory provides a rigorous framework for characterizing the behavior of a system in response to perturbations from an equilibrium point. Instead of solving the system's equations explicitly, we analyze its qualitative properties, such as whether trajectories remain close to, or converge to, an equilibrium. We assume without loss of generality that the equilibrium of interest is at the origin ($\boldsymbol{x}^* = \mathbf{0}$).

Definitions for Stability Analysis

The analysis of stability, particularly via Lyapunov's method, relies on specific classes of functions that characterize properties like positivity and growth.

Definition 2.4.27 (Positive Definite and Radially Unbounded Functions)**.** *Let $V : \mathcal{D} \to \mathbb{R}$ be a continuous function on a domain $\mathcal{D} \subset \mathbb{R}^N$ containing the origin.*

- *V is **positive semi-definite** centered at $\boldsymbol{x}^*$ if $V(\boldsymbol{x}^*) = 0$ and $V(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathcal{D} \setminus \{\boldsymbol{x}^*\}$.*
- *V is **positive definite** centered at $\boldsymbol{x}^*$ if $V(\boldsymbol{x}^*) = 0$ and $V(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{D} \setminus \{\boldsymbol{x}^*\}$.*
- *V is **radially unbounded** if $\mathcal{D} = \mathbb{R}^N$ and $V(\boldsymbol{x}) \to \infty$ as $\|\boldsymbol{x}\| \to \infty$.*

Definition 2.4.28 (Class $\mathcal{K}$ and $\mathcal{K}_\infty$ Functions [117, Definition 4.2])**.** *A function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to be a **class $\mathcal{K}$ function** if it is continuous, strictly increasing and $\alpha(0) = 0$. Moreover, it is of **class $\mathcal{K}_\infty$** if it is a class $\mathcal{K}$ function and $\lim_{r\to+\infty} \alpha(r) = +\infty$.*

Definition 2.4.29 (Class $\mathcal{KL}$ Functions [117, Definition 4.3])**.** *A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to be a **class $\mathcal{KL}$ function** if, for each fixed $s \geq 0$, $\beta(\cdot, s)$ is a class $\mathcal{K}$ function, and for each fixed $r > 0$, $\beta(r, \cdot)$ is strictly decreasing with respect to s and $\lim_{s\to+\infty} \beta(r, s) = 0$.*

[117] *Nonlinear Systems*. Hassan K Khalil. (2002).

Notions of Stability

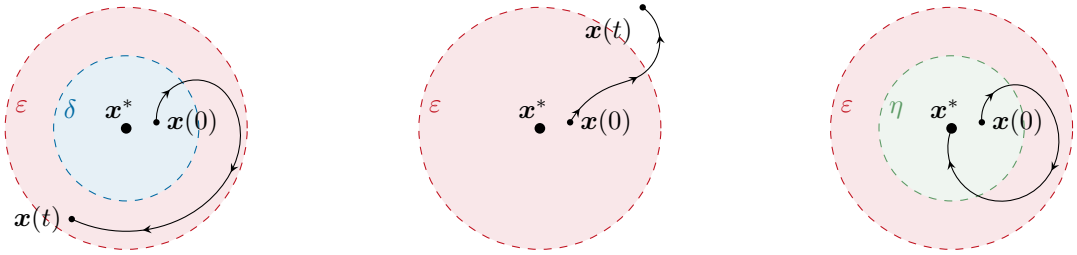
With these core concepts, we can formally define the different types of stability.

Definition 2.4.30 (Stability of an Equilibrium Point [117, Definition 4.1])**.** *Let $\boldsymbol{x}^* \in \mathcal{X}$ be an equilibrium point of the system (2.5). We say that:*

- *$\boldsymbol{x}^*$ is **stable** (or Lyapunov stable) if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that if $\|\boldsymbol{x}(0) - \boldsymbol{x}^*\| < \delta$, then $\|\boldsymbol{x}(t) - \boldsymbol{x}^*\| < \varepsilon$ for all $t \geq 0$. (Trajectories starting nearby remain nearby.)*
- *$\boldsymbol{x}^*$ is **unstable** if it is not stable. (Trajectories starting nearby may move away.)*
- *$\boldsymbol{x}^*$ is **asymptotically stable** if it is stable and, in addition, there exists an $\eta > 0$ such that if $\|\boldsymbol{x}(0) - \boldsymbol{x}^*\| < \eta$, then $\lim_{t\to\infty} \boldsymbol{x}(t) = \boldsymbol{x}^*$. (Trajectories starting nearby converge to $\boldsymbol{x}^*$.)*

*The set of all initial conditions that converge to an asymptotically stable equilibrium point is called its **region of attraction**. If the region of attraction is the entire state space $\mathcal{X}$, the equilibrium is **globally asymptotically stable**.*

[117] *Nonlinear Systems*. Hassan K Khalil. (2002).



(a) Stable (b) Unstable (c) Asymptotically Stable

Figure 2.9: Visual representation of stability notions for an equilibrium point $\boldsymbol{x}^*$. (a) A stable system: all the trajectories starting within δ remain within ε. (b) An unstable system: trajectories starting within ε eventually leave ε. (c) An asymptotically stable system: trajectories starting within η converge to $\boldsymbol{x}^*$.

To further clarify these critical notions of stability, we provide illustrative figures for each case in Figure 2.9.

Definition 2.4.31 (Exponential Stability)**.** *An equilibrium point $\boldsymbol{x}^*$ is **locally exponentially stable** if it is locally asymptotically stable and there exist positive constants c_1, c_2, η such that for all $\boldsymbol{x}(0)$ with $\|\boldsymbol{x}(0) - \boldsymbol{x}^*\| < \eta$:*

$$\|\boldsymbol{x}(t) - \boldsymbol{x}^*\| \leq c_1 \|\boldsymbol{x}(0) - \boldsymbol{x}^*\| e^{-c_2 t}.$$

*If this holds for all $\boldsymbol{x}(0) \in \mathcal{X}$ (i.e., the region of attraction is $\mathcal{X}$), the equilibrium is **globally exponentially stable**. Exponential stability implies a uniform, rapid rate of convergence c_2.*

Analytical Tools for Stability: The Comparison Principle

To rigorously analyze stability, we often need to infer the long-term behavior of a function (e.g., a Lyapunov function $V(\boldsymbol{x}(t))$) from a bound on its derivative ($\dot{V}$). The following analytical tools provide the formal link between a differential inequality and a bound on its solution over time. The most important of these is the Comparison Lemma. But first, introduce the notion of upper right-derivative.

Definition 2.4.32 (Upper Right-Derivative [116, Definition 1.24])**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a function. The **upper right-derivative**, also known as the **Dini derivative** of f at x is defined as:*

$$D^+ f(x) = \limsup_{h \to 0^+} \frac{f(x+h) - f(x)}{h},$$

which could be infinite.

[116] *Fonctions de Lyapunov : Stabilité.* Delphine Bresch-Pietri et al. (2022).

We can now state the Comparison Lemma.

Comparison Lemma ([117, Lemma 3.4]). *Consider the scalar differential equation:*

$$\dot{x}(t) = f(x(t)), \quad x(t_0) = x_0,$$

where f is continuous and locally Lipschitz. Let I be the maximal interval of existence of the solution $x(t)$. Let $v : I \to \mathbb{R}$ be a continuously differentiable function. Suppose that $v(t)$ satisfies the differential inequality:

$$D^+v(t) \leq f(v(t)), \quad \forall t \in I, \quad \text{and} \quad v(t_0) \leq x_0,$$

where D^+ is the upper right-derivative. Then, for as long as both solutions exist, we have:

$$v(t) \leq x(t), \quad \forall t \in I.$$

[117] *Nonlinear Systems*. Hassan K Khalil. (2002).

A direct and frequently used consequence of the Comparison Lemma is Gronwall's inequality, which provides explicit exponential bounds.

Gronwall's Inequality ([122] or [116, Theorem 1.59]). *Let $v(t)$ be a continuously differentiable function satisfying the differential inequality*

$$\dot{v}(t) \leq k\, v(t)$$

for some constant $k \in \mathbb{R}$ and all $t \geq t_0$. Then, for all $t \geq t_0$:

$$v(t) \leq v(t_0)e^{k(t-t_0)}.$$

In particular, if $\dot{v}(t) \leq -c\, v(t)$ for some $c > 0$, then $v(t)$ decays exponentially: $v(t) \leq v(t_0)e^{-c(t-t_0)}$.

[122] "Note on the Derivatives with Respect to a Parameter of the Solutions of a System of Differential Equations". T. H. Gronwall. (1919).
[116] *Fonctions de Lyapunov : Stabilité*. Delphine Bresch-Pietri et al. (2022).

Lyapunov's Direct Method

Lyapunov's direct method is the basic tool of stability analysis, providing a way to prove stability without explicitly solving the system's differential equations. The core idea is to find a scalar, "energy-like" function $V : \mathcal{X} \to \mathbb{R}$ whose properties along the system's trajectories reveal the stability of an equilibrium. For this to work, such a function must satisfy two intuitive conditions:

(i) It must be positive definite, meaning it is zero at the equilibrium and positive everywhere else, thus acting as a generalized measure of the state's distance from equilibrium.

(ii) Its value must be non-increasing along all solution trajectories. This is verified by checking that its time derivative, $\dot{V}(\boldsymbol{x})$, is negative semi-definite (i.e., $\dot{V} \leq 0$). A strict decrease in energy corresponds to a negative definite derivative ($\dot{V} < 0$).

The following definition formalizes these requirements for different types of Lyapunov functions.

Definition 2.4.33 (Lyapunov Function). *Let $\boldsymbol{x}^*$ be an equilibrium point of the system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$. A continuously differentiable function $V : \mathcal{X} \to \mathbb{R}$ is called a:*

- ***Lyapunov function candidate*** *if $V(\boldsymbol{x}^*) = 0$.*
- ***weak Lyapunov function*** *if $V(\boldsymbol{x}^*) = 0$ and $V(\boldsymbol{x}) > 0$ for $\boldsymbol{x} \neq \boldsymbol{x}^*$ in a neighborhood of $\boldsymbol{x}^*$, and its Lie derivative*
$$\dot{V}(\boldsymbol{x}) = \nabla V(\boldsymbol{x}) \cdot \boldsymbol{f}(\boldsymbol{x})$$
is locally negative semi-definite around $\boldsymbol{x}^*$. i.e., $\dot{V}(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x} \in B(\boldsymbol{x}^*, \delta)$ for some $\delta > 0$.
- ***local Lyapunov function*** *if it is a weak Lyapunov function and $\dot{V}(\boldsymbol{x})$ is locally negative definite around $\boldsymbol{x}^*$. i.e., $\dot{V}(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \in B(\boldsymbol{x}^*, \delta)$ for some $\delta > 0$.*
- ***global Lyapunov function*** *if it is a local Lyapunov function, $V(\boldsymbol{x})$ is globally positive definite and radially unbounded. i.e., $V(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{X} \setminus \{\boldsymbol{x}^*\}$ and $\lim_{\|\boldsymbol{x}\| \to \infty} V(\boldsymbol{x}) = \infty$.*

The power of this method is that just the existence of a function satisfying these properties is sufficient to guarantee stability. The following theorems make this connection explicit.

Theorem 2.4.34 (Lyapunov Stability Criteria [117, Theorem 4.1]). *Consider a dynamical system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ with an equilibrium point $\boldsymbol{x}^*$.*

*(i) The equilibrium point $\boldsymbol{x}^*$ is **stable** if there exists a weak Lyapunov function V.*

*(ii) The equilibrium point $\boldsymbol{x}^*$ is **locally asymptotically stable** if there exists a local Lyapunov function V.*

*(iii) The equilibrium point $\boldsymbol{x}^*$ is **globally asymptotically stable** if there exists a global Lyapunov function V.*

*(iv) The equilibrium point $\boldsymbol{x}^*$ is **globally exponentially stable** if there exist class $\mathcal{K}_\infty$ functions α_1, α_2 and a class $\mathcal{K}$ function α_3 such that:*
$$\alpha_1(\|\boldsymbol{x}\|) \leq V(\boldsymbol{x}) \leq \alpha_2(\|\boldsymbol{x}\|) \quad \text{and} \quad \dot{V}(\boldsymbol{x}) \leq -\alpha_3(\|\boldsymbol{x}\|).$$
More specific criteria for exponential stability often involve quadratic bounds:
$$k_1\|\boldsymbol{x}\|^2 \leq V(\boldsymbol{x}) \leq k_2\|\boldsymbol{x}\|^2 \quad \text{and} \quad \dot{V}(\boldsymbol{x}) \leq -k_3\|\boldsymbol{x}\|^2,$$
for some positive constants k_1, k_2, k_3.

[117] *Nonlinear Systems*. Hassan K Khalil. (2002).

2.4.5 Invariance Principle

LaSalle's Principle extends Lyapunov's method, allowing one to conclude asymptotic stability even when $\dot{V}$ is only negative semi-definite (a weak Lyapunov function). This principle is particularly useful for analyzing the stability of complex systems with multiple equilibrium points.

LaSalle's Invariance Principle ([117, Theorem 4.4]). *Let $\Omega \subset D$ be a compact set that is positively invariant with respect to the autonomous system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ where $\boldsymbol{f} : D \to \mathbb{R}^N$ is a locally Lipschitz map from a domain $D \subset \mathbb{R}^N$ into $\mathbb{R}^N$. Let $V : D \to \mathbb{R}$ be a continuously*

differentiable function such that $\dot{V}(x) \leq 0$ for all $x \in \Omega$. Let S be the set of all points in Ω where $\dot{V}(x) = 0$, that is,

$$S = \{\boldsymbol{x} \in \Omega \mid \dot{V}(\boldsymbol{x}) = 0\} \tag{2.7}$$

Let M be the largest invariant set in S. Then every solution starting in Ω approaches M as $t \to \infty$.

[117] *Nonlinear Systems*. Hassan K Khalil. (2002).

2.4.6 Input-to-State Stability

We now generalize the concept of stability to systems with external inputs, $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u})$. Input-to-State Stability (ISS) provides a framework to analyze how the state is bounded by the decaying influence of the initial condition and the persistent influence of the input. This requires a specialized class of comparison functions.

Definition 2.4.35. *Consider a nonlinear system described by the differential equation*

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}(t), \boldsymbol{u}(t))$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the state vector and $\boldsymbol{u}(t) \in \mathbb{R}^m$ is the input vector. Let $\boldsymbol{x}(t)$ denote the solution starting from the initial state $\boldsymbol{x}(0) = \boldsymbol{x}_0$ under the input $\boldsymbol{u}(t)$.

*The system is said to be **Input-to-State Stable (ISS)** if there exist functions $\gamma : [0, \infty) \to [0, \infty)$ and $\beta : [0, \infty) \times [0, \infty) \to [0, \infty)$ such that:*

(i) γ is a $\mathcal{K}$ function. i.e., it is continuous, strictly increasing, and $\gamma(0) = 0$.

(ii) β is a $\mathcal{KL}$ function. i.e., for any fixed $t \geq 0$, $\beta(\cdot, t)$ is a $\mathcal{K}$ function, and for any fixed $r > 0$, $\beta(r, t)$ is decreasing with respect to t and $\lim_{t\to\infty} \beta(r, t) = 0$.

And for any initial state $\boldsymbol{x}_0 \in \mathbb{R}^n$ and any bounded measurable input $\boldsymbol{u} : [0, \infty) \to \mathbb{R}^m$, the solution $\boldsymbol{x}(t)$ exists for all $t \geq 0$ and satisfies the inequality:

$$\|\boldsymbol{x}(t)\| \leq \beta(\|\boldsymbol{x}_0\|, t) + \gamma\big(\sup_{0\leq\tau\leq t} \|\boldsymbol{u}(\tau)\|\big) \tag{2.8}$$

for all $t \geq 0$.

Remark 2.4.3. *This inequality implies two key properties:*

- *If $\boldsymbol{u}(t) \equiv 0$, the system is globally asymptotically stable with respect to the origin, as $\|\boldsymbol{x}(t)\| \leq \beta(\|\boldsymbol{x}_0\|, t) \to 0$ as $t \to \infty$.*
- *If the input is bounded ($\|\boldsymbol{u}(t)\| \leq U$ for some U), the state is ultimately bounded: $\limsup_{t\to\infty} \|\boldsymbol{x}(t)\| \leq \gamma(U)$.*

Moreover, since the system is finite-dimensional, the condition is satisfied with any norm by equivalence of norms.

2.5 Mathematical Tools

In addition to properties of networks and dynamical systems, the analysis of dynamical systems relies on fixed-point theory to establish the existence of equilibria. This section summarizes the key theorems used to guarantee that valid solutions and steady states exist for the algorithms proposed in this thesis.

We begin with the classical Intermediate Value Theorem. While this is a standard result in real analysis, it serves as a fundamental building block for existence proofs involving continuous scalar functions.

Intermediate Value Theorem. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function defined on a closed interval $[a, b]$. For any value y between $f(a)$ and $f(b)$, there exists at least one number $c \in (a, b)$ such that $f(c) = y$.*

For higher-dimensional systems, specifically when dealing with continuous dynamics on bounded sets, we require stronger topological tools. Brouwer's Fixed Point Theorem is the standard result used to prove that a system has at least one equilibrium point. This is required for validating that the control protocols we design have a state to converge to.

Brouwer's Fixed Point Theorem. *[123] Let $f : \mathcal{B} \to \mathcal{B}$ be a continuous function defined on a closed ball $\mathcal{B} \subseteq \mathbb{R}^N$. Then, f has a fixed point in $\mathcal{B}$.*

[123] *Fixed Point Theory*. Andrzej Granas et al. (2003).

Finally, not all systems studied in this thesis rely on continuity. For systems involving discrete states or strictly monotone update rules, order-theoretic properties are often more relevant than topological ones. The Knaster-Tarski theorem guarantees the existence of fixed points for order-preserving mappings. This is particularly useful for analyzing the convergence of algorithms that maintain a specific ordering between agents, such as in monotone consensus problems.

Knaster-Tarski Fixed-Point Theorem. *[123, Theorem 1.1] Let $(L, \leq)$ be a complete lattice, i.e., a partially ordered set in which every subset has both a supremum and an infimum. Let $f : L \to L$ be an order-preserving (or monotone) function, meaning that for all $x, y \in L$, if $x \leq y$, then $f(x) \leq f(y)$.*

Then the set of fixed points of f, denoted $\mathrm{Fix}(f) = \{x \in L \mid f(x) = x\}$, has the following properties:

(i) The set $\mathrm{Fix}(f)$ is non-empty.

(ii) The set $(\mathrm{Fix}(f), \leq)$ is also a complete lattice.

Furthermore, the least fixed point and the greatest fixed point of f exist and are given by:

$$\inf(\mathrm{Fix}(f)) = \inf\{x \in L \mid f(x) \leq x\} \quad \textit{(the least fixed point)}$$
$$\sup(\mathrm{Fix}(f)) = \sup\{x \in L \mid x \leq f(x)\} \quad \textit{(the greatest fixed point)}$$

[123] *Fixed Point Theory*. Andrzej Granas et al. (2003).

Chapter 3

A non-linear opinion dynamics model: Continuous Opinion Signaled Action (COSA)

3.1 Introduction

In this chapter, we introduce the **Continuous Opinion Signaled Action (COSA)** model, a natural generalization of the **Continuous Opinion Discrete Action (CODA)** framework. The COSA model extends CODA by allowing agents to communicate through general signal functions rather than binary actions, thereby capturing a broader spectrum of interaction mechanisms ranging from fully quantized to fully transparent communication.

We begin by formally defining the model and its structural properties, paying particular attention to the role of the communication graph and the influence of the signal function on agent dynamics. We then establish the standing assumptions that will be maintained throughout the remainder of the thesis. These results provide the foundation for subsequent analysis: in the next chapter, COSA will be used to study synchronization phenomena, both full and partial. Later in the thesis, the model will be enriched by coupling it to an external environment, enabling us to analyze how collective behavior is shaped by external influences.

3.2 Continuous Opinion Signaled Action (COSA)

Before introducing the COSA update rule, we first describe the agents and the communication graph that mediates their interactions.

3.2.1 Model Setup: Agents and Communication Graph

We consider a multi-agent system composed of a population of N agents, indexed by the set $\mathcal{V} = \{1, \dots, N\}$. Each agent $i \in \mathcal{V}$ holds a continuous opinion, denoted by $x_i \in [-1, 1]$, where 0 represents a neutral stance and ± 1 represent opposing extremist positions.

The interactions among agents are described by a static, communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. The presence of an edge $(i, j) \in \mathcal{E}$ signifies that agents i can influence j. The set of neighbors of agent i is denoted by $\mathcal{N}_i = \{j \in \mathcal{V} \mid (j, i) \in \mathcal{E}\}$.

The graph topology is encoded by several matrices:

- The **adjacency matrix** $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is a binary matrix with entries $a_{ij} = 1$ if $(i,j) \in \mathcal{E}$ (i.e., agent i influences agent j), and $a_{ij} = 0$ otherwise. We consider unweighted graphs for simplicity, thus $w_{ij} = 1$ for existing edges.
- The **in-degree matrix** $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ is a diagonal matrix whose diagonal entries are the **in-degrees** of the agents, $d_i = |\mathcal{N}_i| = \sum_{j=1}^{N} a_{ji}$. This matrix records the number of incoming connections for each agent.
- The **Laplacian matrix** is defined as $\boldsymbol{L} := \boldsymbol{D} - \boldsymbol{A}$.

These matrices allow us to describe information flow on the network in algebraic terms. In what follows, we will work under a standard structural assumption on the communication graph.

Standing Assumption 3.1. *The communication graph $\mathcal{G}$ is connected and undirected.*

Assumption 3.1 ensures that the communication graph $\mathcal{G}$ is both connected and undirected. These conditions are standard in the consensus and opinion dynamics literature for the following reasons:

- **Connectedness:** If the graph were not connected, then the system would naturally decompose into independent connected components, each evolving according to its own dynamics. In that case, analyzing the global system adds no additional insight beyond analyzing each component separately. Hence, the connectedness assumption allows us to focus on a single, nontrivial collective dynamics.
- **Undirectedness:** By restricting attention to undirected graphs, we assume that influence between two connected agents is bidirectional. This symmetry significantly simplifies the analysis, since the associated Laplacian matrix is symmetric and its spectral properties are well understood. At the same time, this assumption still captures a broad class of realistic social and communication networks, where reciprocal interactions are a reasonable modeling choice.

In summary, this standing assumption makes a balance between analytical tractability and practical relevance, allowing us to capture the essential features of inter-agent communication without unnecessary technical complications.

Notation 3.2.1. *In the following of this thesis, we will denote $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_N$ as the eigenvalues of $\boldsymbol{D}^{-1}\boldsymbol{A}$ ordered in non-decreasing order, and $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$ as the corresponding right eigenvectors. Moreover, we will denote $\mu_1 \le \mu_2 \le \cdots \le \mu_N$ as the eigenvalues of $\boldsymbol{D}^{-1}\boldsymbol{L}$ ordered in non-decreasing order, and $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_N$ as the corresponding right eigenvectors.*

Proposition 3.2.1 (Spectral Properties of $\boldsymbol{D}^{-1}\boldsymbol{A}$ for Symmetric Graphs)**.** *Under Standing Assumption 3.1 (connected, undirected $\mathcal{G}$), let $\boldsymbol{A}_{\mathrm{rw}} := \boldsymbol{D}^{-1}\boldsymbol{A}$ and $\boldsymbol{L}_{\mathrm{rw}} := \boldsymbol{D}^{-1}\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{A}$. Then:*

(i) The eigenvalues of $\boldsymbol{A}_{\mathrm{rw}}$ (and consequently of $\boldsymbol{L}_{\mathrm{rw}}$) are real.

(ii) $\boldsymbol{A}_{\mathrm{rw}}$ and the symmetric matrix $\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ share the same eigenvalues.

(iii) $\boldsymbol{A}_{\mathrm{rw}}$ is irreducible.

(iv) $\lambda_{\max}(\boldsymbol{A}_{\mathrm{rw}}) = 1$ is simple, with right eigenvector $\mathbf{1}$.

(v) The spectrum of $\boldsymbol{L}_{\mathrm{rw}}$ is $\sigma(\boldsymbol{L}_{\mathrm{rw}}) = \{\mu_i = 1 - \lambda_i \mid i = 1, \ldots, N\}$, so $\mu_1 = 0$ and $\mu_N = 1 - \lambda_1$.

(vi) The eigenvalues of $\boldsymbol{A}_{\mathrm{rw}}$ lie in $[-1, 1]$ with $\lambda_{\max} = 1$. Moreover, $-1 \in \sigma(\boldsymbol{A}_{\mathrm{rw}})$ if and only if $\mathcal{G}$ is bipartite. Consequently, $\sigma(\boldsymbol{L}_{\mathrm{rw}}) \subseteq [0, 2]$, and $2 \in \sigma(\boldsymbol{L}_{\mathrm{rw}})$ if and only if $\mathcal{G}$ is bipartite.

(vii) *$\boldsymbol{A}_{\mathrm{rw}}$ is diagonalizable over $\mathbb{R}$; its eigenvectors can be chosen orthonormal for the weighted inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{D}} = \boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{y}$.*
We define the projection $\boldsymbol{Q}_i$ onto the eigenspace of $\boldsymbol{A}_{\mathrm{rw}}$ associated to λ_i. i.e.,

$$\boldsymbol{Q}_i \boldsymbol{x} = \langle \boldsymbol{v}_i, \boldsymbol{x} \rangle_{\boldsymbol{D}} \boldsymbol{v}_i \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^N.$$

Then, $\boldsymbol{Q}_i$ is self-adjoint and $\boldsymbol{Q}_i \boldsymbol{Q}_j = \delta_{ij} \boldsymbol{Q}_i$.

(viii) *If $\mathcal{G}$ is d-regular, then $\boldsymbol{A}_{\mathrm{rw}} = \frac{1}{d}\boldsymbol{A}$ and $\sigma(\boldsymbol{A}_{\mathrm{rw}}) = \{\lambda_{\boldsymbol{A}}/d \mid \lambda_{\boldsymbol{A}} \in \sigma(\boldsymbol{A})\}$; equivalently, $\boldsymbol{L}_{\mathrm{rw}} = \boldsymbol{I} - \frac{1}{d}\boldsymbol{A}$ and $\sigma(\boldsymbol{L}_{\mathrm{rw}}) = \{1 - \lambda_{\boldsymbol{A}}/d \mid \lambda_{\boldsymbol{A}} \in \sigma(\boldsymbol{A})\}$.*

Proof. (i)–(ii) By Proposition 2.2.16, $\boldsymbol{D}^{-1}\boldsymbol{A}$ is similar to the symmetric matrix $\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$, and therefore has real eigenvalues and shares the same spectrum.

(iii)–(iv) By Perron-Frobenius Theorem, since the graph is connected (Standing Assumption 3.1), the matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$ is irreducible. Additionally, $\lambda = 1$ is an eigenvalue with right eigenvector $\mathbf{1}$ and is simple.

(v)–(viii) Proposition 2.2.15 shows that the spectrum of $\boldsymbol{L}_{\mathrm{rw}}$ is obtained from that of $\boldsymbol{A}_{\mathrm{rw}}$ by the relation $\mu_i = 1 - \lambda_i$. In case of a d-regular graph, $\boldsymbol{A}_{\mathrm{rw}} = \frac{1}{d}\boldsymbol{A}$, so $\sigma(\boldsymbol{A}_{\mathrm{rw}}) = \{\lambda_{\boldsymbol{A}}/d : \lambda_{\boldsymbol{A}} \in \sigma(\boldsymbol{A})\}$.

(vi) By Proposition 2.2.16, since $\boldsymbol{A}_{\mathrm{rw}}$ is similar to a symmetric matrix, its eigenvalues lie in $[-1, 1]$. Proposition 2.2.14 shows that $-1 \in \sigma(\boldsymbol{A}_{\mathrm{rw}})$ if and only if $\mathcal{G}$ is bipartite, which via Proposition 2.2.15 is equivalent to $2 \in \sigma(\boldsymbol{L}_{\mathrm{rw}})$.

(vii) By Proposition 2.2.16, $\boldsymbol{A}_{\mathrm{rw}}$ is diagonalizable over $\mathbb{R}$ and its eigenvectors can be chosen to form a $\boldsymbol{D}$-orthonormal basis. □

The spectral facts above explain how information propagates on $\mathcal{G}$: $\boldsymbol{A}_{\mathrm{rw}}$ averages neighbor values, its simple eigenvalue 1 (with eigenvector $\mathbf{1}$) singles out the consensus direction, and the remaining spectrum governs the decay of disagreement.

Having established the spectral properties of the communication graph, we now turn to the modeling of opinion dynamics. Our starting point is the well-known CODA model, which provides the foundation upon which COSA is built on.

3.2.2 From CODA to COSA

The starting point of our framework is the **Continuous Opinion Discrete Action (CODA)** model introduced by [101] as a stochastic model that was later proved to be equivalent numerically to the differential model proposed in [67]. In this model, agents hold continuous opinions $x_i(k) \in [0, 1]$, but at every time step, agent i observes only the discrete actions of its neighbors $\mathcal{N}_i$.

The CODA update with step size 1 rule is then

$$x_i(k+1) = x_i(k) + x_i(k)(1 - x_i(k))\Big(\frac{1}{|\mathcal{N}_i|}\sum_{j \in \mathcal{N}_i} q_j(k) - x_i(k)\Big), \tag{CODA}$$

where the action of agent j is defined as

$$q_j(k) = \begin{cases} 1 & \text{if } x_j(k) \geq 1/2, \\ 0 & \text{if } x_j(k) < 1/2. \end{cases} \tag{3.1}$$

[101] "Continuous Opinions and Discrete Actions in Opinion Dynamics Problems". André C. R. Martins. (2008).
[67] "Continuous Opinions and Discrete Actions in Social Networks". N. R. Chowdhury et al. (2016).

The multiplicative factor $x_i(1 - x_i)$ in (CODA) acts as a *stubbornness term*; mathematically, it is a state-dependent gain on the agent's update rule, which vanishes as opinions approach the extremes, making agents less responsive to influence.

For our purposes, however, it is more convenient to recenter the opinion space symmetrically around zero. By applying the transformation $x \mapsto 2x - 1$, the CODA model is equivalently expressed on the domain $[-1, 1]$, where neutrality corresponds to 0 and opposing camps are represented at ± 1. In this rescaled formulation, the quantized action of an agent reduces to

$$q_j(k) = \text{sign}(x_j(k)),$$

making the symmetric structure of the model explicit and more natural for analyzing polarization dynamics. The action $q_j(k)$ can be seen as a *signal* sent by agent j to agent i at time k. i.e. $q_j(k) = s(x_j(k))$ where $s(x) = \text{sign}(x)$ is the sign function defined as $s(x) = 1$ if $x > 0$, $s(x) = -1$ if $x < 0$, and $s(0) = 0$.

This observation motivates the **Continuous Opinion Signaled Action (COSA)** model, which generalizes CODA by replacing the binary action with a flexible, possibly nonlinear, *signal function* $s : [-1, 1] \to [-1, 1]$. This function is assumed to be non-decreasing and maps an agent's internal opinion to the information it communicates to its neighbors. The dynamics for agent i then become

$$\dot{x}_i(t) = \big(1 - x_i(t)^2\big)^{\alpha} \Big(\frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} s(x_j(t)) - x_i(t) \Big), \tag{3.2}$$

with a discrete-time counterpart with step size h:

$$x_i(k+1) = x_i(k) + h \left(\big(1 - x_i(k)^2\big)^{\alpha} \Big(\frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} s(x_j(k)) - x_i(k) \Big) \right). \tag{3.3}$$

The parameter $\alpha \geq 0$ provides a modular way to capture *stubbornness*:

- If $\alpha = 0$, it eliminates stubbornness altogether, leaving only a uniform linear relaxation toward the average signal. This yields a simpler, analytically tractable system in which the nonlinear effects of $s(x)$ can be studied in isolation.
- If $\alpha = 1$, the model reproduces the opinion-dependent inertia of the original CODA model, where opinions near ± 1 are less prone to change.
- For larger values of α it accentuates the stubbornness, increasingly slowing dynamics near the extremes.

This general formulation unifies several well-known models. When $s(x)$ approaches a step function, CODA is recovered; when $s(x) = x$ and $\alpha = 0$, the model reduces to linear consensus with symmetric bounded opinions. In this sense, COSA interpolates between discontinuous and fully transparent communication, while the parameter α tunes the extent of endogenous stubbornness.

For the following chapters, however, we will focus on the case $\alpha = 0$, i.e., COSA *without stubbornness*. This choice is primarily motivated by traceability: it removes the additional nonlinearities associated with opinion-dependent inertia, thereby allowing us to isolate and rigorously analyze the role of the signal function $s(x)$ in shaping emergent collective behavior.

We are now ready to present the COSA model formally.

3.2.3 The COSA model

The state of agent $i \in \mathcal{V}$ is denoted by $x_i \in [-1, 1]$ and evolves according to

$$\dot{x}_i(t) = \frac{1}{d_i} \sum_{j=1}^{N} a_{ij} s(x_j(t)) - x_i(t), \tag{agent COSA}$$

where $s : [-1, 1] \to [-1, 1]$ is a **common non-decreasing signal function**. The discrete-time version with step size h is

$$x_i(k+1) = x_i(k) + h \left(\frac{1}{d_i} \Big(\sum_{j=1}^{N} a_{ij} s(x_j(k)) \Big) - x_i(k) \right). \tag{agent COSA-DT}$$

These dynamics capture imperfect communication among agents through the signal function $s(x)$, which models how opinions are perceived and transmitted. Typical choices include the linear map $s(x) = x$ (unbiased estimation), affine maps $s(x) = ax + b$ (biased estimation), and nonlinear functions such as $s(x) = \tanh(x)$, which are motivated by biological or neuroscientific mechanisms [92, 96].

The framework also accommodates quantized communication. Discontinuous quantization schemes [67, 101, 105] can be approximated by continuous piecewise affine functions, preserving the qualitative behavior: synchronization, when it occurs, settles at fixed points of $s(x)$. In practical terms, $s(x)$ can be interpreted as the expectation of a quantized signal, possibly smoothed by small noise. Thus, (agent COSA) provides a versatile and analytically tractable model bridging idealized linear consensus and more realistic nonlinear or discontinuous interactions.

For collective dynamics, we write

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{D}^{-1} \boldsymbol{A} \boldsymbol{s}(\boldsymbol{x}(t)) - \boldsymbol{x}(t) := \boldsymbol{f}(\boldsymbol{x}(t)), \tag{COSA}$$

and its discrete-time counterpart with step size h:

$$\boldsymbol{x}(k+1) = \boldsymbol{x}(k) + h \left(\boldsymbol{D}^{-1} \boldsymbol{A} \boldsymbol{s}(\boldsymbol{x}(k)) - \boldsymbol{x}(k) \right). \tag{COSA-DT}$$

where $\boldsymbol{x} = (x_1, \ldots, x_N)^\top \in \mathcal{X} := [-1, 1]^N$ and $\boldsymbol{s}(\boldsymbol{x}) = (s(x_1), \ldots, s(x_N))^\top \in \mathcal{X}$.

It is important to emphasize the role of the normalized adjacency matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$ (also known as the *random-walk adjacency matrix*) in (COSA). Using this normalized form, rather than the raw adjacency matrix $\boldsymbol{A}$, has two main advantages:

- First, it ensures that the dynamics remain confined to the bounded domain $[-1, 1]^N$, so that the signal function s is always well-defined.
- Second, it equalizes the interaction speed across agents regardless of their degree.
- In addition, since $\boldsymbol{D}^{-1}\boldsymbol{A}$ is row-stochastic, the Perron-Frobenius Theorem applies: for connected graphs it has a simple eigenvalue 1 with eigenvector $\mathbf{1}$, while all other eigenvalues lie strictly inside the unit circle.

This means that normalization changes only the timescale of convergence, not the structure of the synchronization equilibria.

[92] "Multiagent Decision-Making Dynamics Inspired by Honeybees". Rebecca Gray et al. (2018).
[96] "Nonlinear Opinion Dynamics With Tunable Sensitivity". Anastasia Bizyaeva et al. (2023).
[105] "Consensus and Disagreement". Francesca Ceragioli et al. (2018).

Remark 3.2.1. *In the special case $s(x) = x$, the dynamics reduce to the normalized linear consensus model:*

$$\dot{\boldsymbol{x}} = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x} = -\boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{x}. \tag{3.4}$$

It is worth noting that the dynamics (COSA) admit several equivalent formulations, each highlighting different structural aspects:

$$\begin{aligned}\dot{\boldsymbol{x}} &= \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x} \\ &= -\boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{x} + \boldsymbol{D}^{-1}\boldsymbol{A}\big(\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}\big) \qquad (3.5)\\ &= -\boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}) + \boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}. \qquad (3.6)\end{aligned}$$

The original form (COSA) most directly reflects the interaction between agents and is the natural extension of the (CODA) to the continuous-time setting.

The second form (3.5) shows that (COSA) can be interpreted as the linear consensus dynamics perturbed by the nonlinear discrepancy between agents' states and their transmitted signals. At the agent level, this reads

$$\dot{x}_i = \underbrace{-\frac{1}{d_i}\sum_{j=1}^{N} a_{ij}(x_i - x_j)}_{\text{linear consensus over states}} + \underbrace{\frac{1}{d_i}\sum_{j=1}^{N} a_{ij}\big(s(x_j) - x_j\big)}_{\text{nonlinear perturbation from neighbors}}, \tag{3.7}$$

which separates the pure consensus mechanism from the nonlinear correction induced by the signal function.

Finally, the third form (3.6) emphasizes consensus directly over the signals, together with an individual-level perturbation:

$$\dot{x}_i = \underbrace{-\frac{1}{d_i}\sum_{j=1}^{N} a_{ij}\big(s(x_i) - s(x_j)\big)}_{\text{consensus over signals}} + \underbrace{\big(s(x_i) - x_i\big)}_{\text{signal–state mismatch at agent } i}. \tag{3.8}$$

The alternative formulations above highlight that COSA can be understood either as a perturbed consensus process or as consensus directly over transmitted signals.

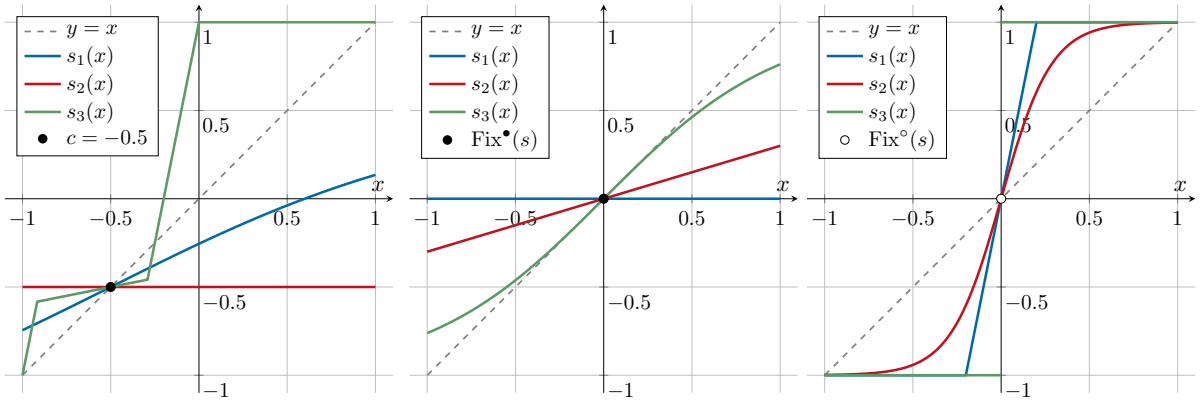
We next examine the properties of the signal function s, which plays the central role in mediating how opinions are perceived and communicated.

3.2.4 Signal Function and Communication Model

Let us introduce the notion of underestimation and overestimation for the communication function s.

In many social and networked settings, agents do not receive neighbors' *states* directly but only *signals* derived from them, often coarse, quantized, or otherwise filtered. This motivates starting from a classical quantized model (CODA) and then broadening it to a signal-based formulation (COSA) that cleanly separates the roles of the graph and of the communication nonlinearity.

To characterize the properties of the signal function, we introduce the following definitions:



(a) Stable estimations of $c = -0.5$. (b) Underestimations (c) Overestimations

Figure 3.1: Examples of different signal functions $s(x)$ classified as stable estimations, underestimations, and overestimations. An understimation can be transformed into a stable estimation by translating and shifting the function $s(x)$ by c. i.e., the function $\tilde{s}(x) = s(x - c) - c$ for c to be a stable estimation of c.

Definition 3.2.2. *A signal function s is said to be:*

- *a **stable estimation** around $c \in \mathrm{Fix}(s)$ if there exists a neighborhood I of c where for all $x \in I$, $(x - c)(s(x) - x) \leq 0$.*
- *an **unstable estimation** around $c \in \mathrm{Fix}(s)$ if there exists a neighborhood I of c where for all $x \in I$, $(x - c)(s(x) - x) > 0$.*
- *an **underestimation** if $x(s(x) - x) \leq 0$, for all $x \in [-1, 1]$. i.e., it is a stable estimation around 0 for the interval $[-1, 1]$.*
- *an **overestimation** if $x(s(x) - x) > 0$, for all $x \in (-1, 1) \setminus \{0\}$. i.e., it is a unstable estimation around 0 for the interval $(-1, 1)$.*

Those definitions are illustrated in Figure 3.1.

To elaborate, a function is an underestimation (resp. overestimation) at a given state x if the processed signal $s(x)$ is closer to (resp. further from) the origin than x itself, relative to the sign of x. This characterizes how the nonlinearity biases the perceived state.

These properties of the signal function s are instrumental in understanding its qualitative behavior. Crucially, the equilibria of our multi-agent system are linked to the *fixed points* of this signal function $\mathrm{Fix}(s)$, where an agent's state matches the signal it generates. We therefore proceed to classify these fixed points based on their local stability characteristics.

Definition 3.2.3. *A fixed point $c \in \mathrm{Fix}(s)$ is classified based on its one-sided stability:*

- *It is **left-stable** (resp. **right-stable**) if there exists a neighborhood I of c such that $(x - c)(s(x) - x) \leq 0$ for all $x \in I$ with $x < c$ (resp. $x > c$).*
- *It is **left-unstable** (resp. **right-unstable**) if the strict inequality $(x - c)(s(x) - x) > 0$ holds for all $x \in I$ with $x < c$ (resp. $x > c$).*

*Based on these properties, we define a fixed point as **stable** if it is both left- and right-stable, **unstable** if it is both left- and right-unstable, and **semi-stable** otherwise.*

Remark 3.2.2. *For example, one can notice that 0 is an unstable fixed point for the sgn () function.*

The local stability defined in Definition 3.2.3 captures a dichotomy. A *stable fixed point* acts as a local attractor: for a state x slightly deviating from c, the dynamics generate a restoring force that pushes x back towards c. In contrast, an *unstable* or *semi-stable fixed point* is locally repelling from at least one side. As we will prove, this simple one-dimensional classification is the key determinant of stability for the full N-dimensional synchronized states.

Based on this property, we partition the set of all fixed points, $\mathrm{Fix}(s)$, into the set of stable fixed points, which we denote by $\mathrm{Fix}^{\bullet}(s)$, and the set of unstable or semi-stable ones, $\mathrm{Fix}^{\circ}(s)$.

$$\mathrm{Fix}(s) = \mathrm{Fix}^{\bullet}(s) \cup \mathrm{Fix}^{\circ}(s).$$

For a more granular analysis, we will also need to identify the specific sources of instability. We define $\mathrm{Fix}^{\circ}_{\mathrm{L}}(s)$ as the set of all fixed points that are *unstable from the left*, and $\mathrm{Fix}^{\circ}_{\mathrm{R}}(s)$ as the set of those that are *unstable from the right*. Note that a fully unstable fixed point belongs to both of these subsets.

$$\mathrm{Fix}^{\circ}(s) = \mathrm{Fix}^{\circ}_{\mathrm{L}}(s) \cup \mathrm{Fix}^{\circ}_{\mathrm{R}}(s).$$

We can now prove the existence of a stable fixed point for any non-decreasing signal function.

Lemma 3.2.4. *Any non-decreasing signal function $s : [-1, 1] \to [-1, 1]$ must have at least one stable fixed point. i.e., $\mathrm{Fix}^{\bullet}(s) \neq \emptyset$.*

Proof. The interval $[-1, 1]$ with the usual ordering is a complete lattice. Since s is a non-decreasing function from this complete lattice to itself, the Knaster-Tarski Fixed-Point Theorem guarantees that the set of fixed points, $\mathrm{Fix}(s)$, is non-empty. Furthermore, the set $\mathrm{Fix}(s)$ itself forms a complete lattice, which implies it is a closed set and contains its own supremum and infimum. Let $c_{\min} = \inf \mathrm{Fix}(s)$ and $c_{\max} = \sup \mathrm{Fix}(s)$. These are the smallest and largest fixed points of s.

We establish two key properties about the behavior of s outside the interval $[c_{\min}, c_{\max}]$.

(i) For any $x > c_{\max}$, we must have $s(x) \leq x$. Indeed, assume for contradiction that there exists an $x_0 \in (c_{\max}, 1]$ with $s(x_0) > x_0$. Consider the interval $J = [x_0, 1]$. For any $y \in J$, we have $y \geq x_0$, and since s is non-decreasing, $s(y) \geq s(x_0) > x_0$. Thus, s maps the complete lattice J into itself. By the Knaster-Tarski theorem, s must have a fixed point $c' \in J$. This implies $c' \geq x_0 > c_{\max}$, which contradicts the definition of $c_{\max}$ as the largest fixed point. The assumption must be false, so $s(x) \leq x$ for all $x > c_{\max}$.

(ii) For any $x < c_{\min}$, we must have $s(x) \geq x$. This can be proved by symmetrical arguments.

We now show that this structure forces at least one fixed point to be stable. Let us assume for contradiction that all fixed points are unstable. Let the fixed points be indexed by $c_i \in \mathrm{Fix}(s)$.

Consider the largest fixed point, $c_{\max}$. From property (i), for any $x > c_{\max}$, we have $x - c_{\max} > 0$ and $s(x) - x \leq 0$, so their product is non-positive. The stability condition holds on the right side of $c_{\max}$. For $c_{\max}$ to be unstable, the violation must occur on its left side. This means that for any $\delta > 0$, there must exist an $x \in (c_{\max} - \delta, c_{\max})$ such that $(x - c_{\max})(s(x) - x) > 0$. Since $x - c_{\max} < 0$, this requires $s(x) < x$.

Let $c_{\text{prev}} = \sup\{c \in \mathrm{Fix}(s) \mid c < c_{\max}\}$ be the fixed point immediately preceding $c_{\max}$ (if one exists). The instability of $c_{\max}$ implies the existence of points $x \in (c_{\text{prev}}, c_{\max})$ where $s(x) < x$. If such a point exists, we can show that $s(y) < y$ for all $y \in (x, c_{\max})$, as $s(y) \geq y$ would imply a fixed point between y and $c_{\max}$. Therefore, the instability of $c_{\max}$ implies $s(x) < x$ for all x in some interval $(c_{\text{prev}}, c_{\max})$.

Now consider this preceding fixed point, c_{prev}. For any $x \in (c_{\text{prev}}, c_{\max})$, the stability product is $(x - c_{\text{prev}})(s(x) - x)$. Since $x > c_{\text{prev}}$ and we have just shown $s(x) < x$, the product is negative.

Thus, the stability condition for c_{prev} holds on its right side. If c_{prev} is also unstable, the violation must occur on its left side, which in turn implies $s(x) < x$ on the interval to the left of c_{prev}.

This argument creates a chain reaction: if a fixed point c_i is unstable, it must be because $s(x) < x$ on the interval (c_{i-1}, c_i), which ensures that the stability condition holds on the right side of c_{i-1}. This means that if all fixed points $c_{\max}, c_{N-1}, \dots, c_2$ are unstable, it is because their instability is passed down from right to left.

Finally, consider the smallest fixed point, $c_{\min}$. If all other fixed points are unstable

However, property (ii) states that for all $x < c_{\min}$, we must have $s(x) \geq x$. This implies $(x - c_{\min})(s(x) - x) \leq 0$. The stability condition for $c_{\min}$ is always satisfied on its left side. This is a contradiction. The assumption that all fixed points are unstable cannot hold.

Therefore, there must exist at least one stable fixed point. □

The above lemma establishes not only the existence of a stable fixed point for any continuous non-decreasing signal function but for any discontinuous one as well.

In the sequel, we will restrict our attention to signal functions s satisfying a few structural properties. These assumptions are weak from a modeling standpoint but crucial from a mathematical one: they guarantee that the induced dynamics can be treated within the framework of dynamical systems introduced in Chapter Networked Systems Preliminaries, especially in Section 2.4. In particular, they ensure that the vector field is well defined, that trajectories exist and are unique (by Theorem 2.4.3), and that we can apply the theory of monotone and cooperative systems developed there.

Thoughout the sequel, we will assume the following standing assumption.

Standing Assumption 3.2. *The signal function $s : [-1, 1] \to [-1, 1]$ is:*

(i) non-decreasing*: $s(x) \leq s(y)$ whenever $x \leq y$,*

(ii) K-Lipschitz continuous *with $K > 0$:*

$$|s(x) - s(y)| \leq K|x - y| \quad \textit{for all } x, y \in [-1, 1],$$

so that the vector field is Lipschitz, guaranteeing existence and uniqueness of solutions.

The monotonicity assumption reflects the natural idea that stronger opinions should not produce weaker signals or actions, while Lipschitz continuity ensures the well-posedness of the dynamics. Together, these conditions make a balance: they are weak enough to cover a broad class of signal functions of interest, yet strong enough to put the system in a highly structured and predictable dynamical behavior, as we will now demonstrate.

Remark 3.2.3. *We will sometimes relax the Lipschitz continuity assumption and consider signal functions that are only non-decreasing. This will allow us to study the behavior of the system in a more general setting.*

3.3 Trajectory Behavior: Invariance and Cooperation

Having formally defined the (COSA) model and its core assumptions, we now turn to analyzing its collective dynamics. A natural first question is whether the system is well-behaved: do trajectories remain bounded, and can they exhibit complex non-convergent patterns like limit cycles or chaos?

This section establishes two foundational properties that provide an answer. We will first prove that the state space is forward-invariant, ensuring all trajectories are bounded. We will then demonstrate that the system is inherently **cooperative** (Definition 2.4.14), a powerful structural property that precludes oscillatory behavior.

3.3.1 Invariance of the state space $\mathcal{X}$

As recalled in Definition 2.4.6, a set $\mathcal{D} \subseteq \mathcal{X}$ is *forward invariant* for (COSA) if every trajectory starting in $\mathcal{D}$ remains in $\mathcal{D}$ for all future time. Our first step is to ensure that trajectories never leave the natural state domain. Establishing the invariance of the hypercube $\mathcal{X} = [-1, 1]^N$ guarantees the boundedness of solutions and the well-posedness of the model on $\mathcal{X}$.

Proposition 3.3.1. *The set $\mathcal{X} = [-1, 1]^N$ is forward invariant for the dynamics (COSA). i.e., if $\boldsymbol{x}(0) \in \mathcal{X}$, then $\boldsymbol{x}(t) \in \mathcal{X}$ for all $t \geq 0$.*

Proof. Since the vector field, (COSA) is Lipschitz-continuous and admit unique solutions starting from any initial condition, we can then apply Nagumo's Theorem to ensure that the solution remains in $\mathcal{X}$.

From Nagumo's Theorem, an analysis at the set's boundary is sufficient to ensure that the solution initialised inside of $\mathcal{X}$ remains inside of $\mathcal{X}$ for all $t \geq 0$. Then, for any $\boldsymbol{x} \in \mathcal{X}$ such that $x_i = 1$ for some $i \in \mathcal{V}$, one has

$$\dot{x}_i = \sum_{j=1}^{N} a_{ij} s(x_j) - x_i = \sum_{j=1}^{N} a_{ij} s(x_j) - 1 \leq \sum_{j=1}^{N} a_{ij} s(1) - 1 = 0.$$

Similarly, for any $\boldsymbol{x} \in \mathcal{X}$ such that $x_i = -1$ for some $i \in \mathcal{V}$, one has

$$\dot{x}_i = \sum_{j=1}^{N} a_{ij} s(x_j) - x_i = \sum_{j=1}^{N} a_{ij} s(x_j) + 1 \geq \sum_{j=1}^{N} a_{ij} s(-1) + 1 = 0.$$

Thus, for any $\boldsymbol{x} \in \partial\mathcal{X}$ (i.e., the boundary of $\mathcal{X}$, i.e., the set of $\boldsymbol{x} \in \mathcal{X}$ such that $x_i = 1$ or $x_i = -1$ for some $i \in \mathcal{V}$), one has $\dot{x}_i(t) \leq 0$ or $\dot{x}_i(t) \geq 0$. Therefore, by Nagumo's Theorem, the solution remains in $\mathcal{X}$ for all $t \geq 0$. □

This proposition confirms that the dynamics are well-defined on $\mathcal{X}$, justifying its choice as the state space. Moreover, this result ensures the boundedness of all trajectories, which is a critical prerequisite for analyzing their long-term behavior and convergence.

3.3.2 Cooperative Properties and Guaranteed Convergence

With boundedness established, we now address the question of the behavior of the dynamics. Does it exhibit complex non-convergent behaviors such as limit cycles or chaos? A priori, a bounded nonlinear system could still exhibit such behaviors. However, the (COSA) model is shielded from such complexities by a powerful structural property: it is inherently **cooperative**.

Intuitively, this means that the influence exerted by any agent j on its neighbors is always positive: an increase in agent j's state, x_j, can only ever encourage its neighbors' states to increase (or remain unchanged), never to decrease. This cooperative principle is a direct consequence of two key structural elements of our model:

- the non-negative nature of the interaction weights in the adjacency matrix ($a_{ij} \geq 0$), reflecting an absence of antagonistic interactions;
- the non-decreasing property of the signal function s, as stipulated in Standing Assumption 3.2, which ensures that a higher input state always translates to a higher or equal perceived signal.

This unidirectional, positive influence has strong implications for the behavior and long-term predictability of the system.

We now formally establish the cooperative properties of the dynamics (COSA).

Proposition 3.3.2. *The dynamics defined by* (COSA) *is cooperative, inducing a monotone flow on the invariant hypercube $\mathcal{X} = [-1, 1]^N$. Specifically, for any two initial conditions $\boldsymbol{x}(0), \boldsymbol{y}(0) \in \mathcal{X}$ such that $\boldsymbol{x}(0) \leq \boldsymbol{y}(0)$ (component-wise), their corresponding trajectories satisfy $\boldsymbol{x}(t) \leq \boldsymbol{y}(t)$ for all $t \geq 0$.*

Proof. To demonstrate that the dynamical system (COSA) induces a monotone flow, from Proposition 2.4.16 we need to show that its vector field $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}$ satisfies the quasi-monotonicity condition (2.6).

Let us consider two state vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ that satisfy the conditions $\boldsymbol{x} \leq \boldsymbol{y}$ and $x_i = y_i$ for a specific index i. We now examine the difference between the i-th components of the vector field at $\boldsymbol{y}$ and $\boldsymbol{x}$:

$$f_i(\boldsymbol{y}) - f_i(\boldsymbol{x}) = \frac{1}{d_i} \sum_{j=1}^{N} a_{ij} \left(s(y_j) - s(x_j)\right) - (y_i - x_i)$$

Since we have the condition $x_i = y_i$, the last term vanishes and for all $i, j \in \mathcal{V}$, $d_i > 0$, $a_{ij} \geq 0$ and s is non-decreasing. Therefore, $f_i(\boldsymbol{x}) \leq f_i(\boldsymbol{y})$. Since this would hold for any component $i \in \mathcal{V}$ such that $x_i = y_i$, we conclude that the vector field $\boldsymbol{f}$ satisfies the quasimonotonicity condition (2.6). □

Proposition 3.3.2 is more than a technical statement. It places strong restrictions on the possible behaviors of the system. Because the flow is now proven to be cooperative, trajectories cannot display oscillatory or chaotic dynamics, which ensures a significant degree of predictability and a much simpler analysis. In particular, the absence of attracting cycles is a powerful simplification as given by the following corollary.

Corollary 3.3.3. *The dynamics* (COSA) *does not admit limit cycles.*

Proof. From Proposition 3.3.2, the dynamics is cooperative. Then, Proposition 2.4.18 ensures that the dynamics does not admit limit cycles. □

The absence of attracting cycles ensures that for almost every initial condition, the system's trajectory will eventually converge to a point within the set of equilibria, $\mathrm{Fix}(\boldsymbol{f})$, which we will now define and classify in Section 3.4. This completely changes our analytical focus since now the central problem is no longer *if* the system settles down, but rather *where* within the set of equilibria it settles.

Under stronger assumptions on the signal function s, one can refine the proof of Proposition 3.3.2 to establish that the dynamics (COSA) is not only cooperative but also generates a *strongly monotone flow*. This stronger property leads to even more robust convergence guarantees, ensuring convergence to a single equilibrium point for all trajectories.

Proposition 3.3.4 (Strongly Monotone Flow for Strictly Increasing Signal Function)**.** *Let s be a nondecreasing Lipschitz continuous signal function. If s is strictly increasing, then the dynamics* (COSA) *is strongly monotone. Moreover, every trajectory of the dynamics* (COSA) *converges to an equilibrium point of* (COSA).

Proof. We need to show that the flow ϕ is strongly monotone under the assumption that s is strictly increasing. Indeed, by Proposition 2.4.24, eventual strong monotonicity implies that the flow is Strongly Order-Preserving (Definition 2.4.23), and then Theorem 2.4.26 ensures that omega-limit sets of bounded trajectories reduce to equilibria.

The strong monotonicity of the dynamics (COSA) follows from Theorem 2.4.22. If s is strictly increasing, then the dependency graph $\mathcal{G}_{\boldsymbol{f}}(\boldsymbol{x})$ (as in Definition 2.4.21) of (COSA) coincides with the interaction graph $\mathcal{G}$, which is strongly connected by assumption. To see this, fix $i \in \mathcal{V}$ and $j \in \mathcal{N}_i$. Consider the function

$$\tilde{f}_i(\xi) \coloneqq f_i(x_1, \dots, x_{j-1}, \xi, x_{j+1}, \dots, x_N).$$

For any interval $[a, b] \subseteq [-1, 1]$ with $a < b$, we compute

$$\begin{aligned}\tilde{f}_i(b) - \tilde{f}_i(a) &= \frac{1}{d_i}\left(\sum_{k=1,k\neq j}^{N} a_{ik}s(x_k) + a_{ij}s(b)\right) - x_i - \left(\frac{1}{d_i}\left(\sum_{k=1,k\neq j}^{N} a_{ik}s(x_k) + a_{ij}s(a)\right) - x_i\right) \\ &= \frac{a_{ij}}{d_i}\left(s(b) - s(a)\right) > 0,\end{aligned}$$

since s is strictly increasing and $a_{ij} > 0$. This shows that $\tilde{f}_i$ is strictly increasing on $[a, b]$, implying that f_i is never independent of x_j on this interval. Since this holds for all $j \in \mathcal{N}_i$, we conclude that $\mathcal{G}_{\boldsymbol{f}}(\boldsymbol{x})$ is strongly connected. We, then, apply Theorem 2.4.22 to ensure that the flow ϕ is strongly monotone. □

The condition of strict monotonicity for s in Proposition 3.3.4 provides a sufficient condition for guaranteeing convergence to a single equilibrium point for all trajectories.

Remark 3.3.1. *If s is only non-decreasing (and not strictly increasing), while attracting cycles are still ruled out by Proposition 3.3.2, the results is not proven for all trajectories, but only for almost all trajectories. This does not means that there could be trajectories that converge to something else than an equilibrium. But, unfortunatly, the tools of strong monotonicity alone does not allows us to prove it in this particular case and remains an open problem. One can infer that it is likely that the system will converge to an equilibrium even in this particular case due to the symmetry of interactions between agents. In the case of directed graphs, example of attracting cycles can be exhibited see [106].*

Collectively, these results provide a crucial guarantee: for almost all initial conditions, trajectories converge to the set of the system's equilibria. This finding a complete shifts our analytical focus for the remainder of the part. The central problem is no longer *if* the system settles down, but rather *where* within the set of equilibria it settles. This validates an equilibrium-centric approach and motivates the subsequent detailed characterization of these steady states, to which we proceed next.

3.4 Characterization of System Equilibria

The preceding analysis provides a powerful guarantee: the cooperative nature of the (COSA) dynamics ensures that trajectories do not exhibit complex oscillatory behavior and, for almost all initial conditions, converge to the set of equilibria. This allows us to shift our focus from the transient dynamics to the properties of these final states.

An equilibrium point $\boldsymbol{x}^* \in \mathcal{X}$ is, by definition, a state satisfying $\dot{\boldsymbol{x}} = \mathbf{0}$. From the system dynamics (COSA), this directly implies:

$$\boldsymbol{x}^* = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}^*). \tag{3.9}$$

Component-wise, this fundamental condition states that for every agent $i \in \mathcal{V}$, its equilibrium state x_i^* must precisely balance its own value with the weighted average of the signals received from its neighbors:

$$x_i^* = \frac{1}{d_i}\sum_{j\in\mathcal{N}_i} s(x_j^*). \tag{3.10}$$

Beyond this per-agent condition, a global constraint on all equilibria can be derived from the spectral properties of the network interaction matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$. This condition provides a collective balance property that every equilibrium must satisfy, reflecting a strong conservation law embedded in the system's structure.

Lemma 3.4.1 (Global Equilibrium Condition)**.** *Let $\boldsymbol{x}^*$ be an equilibrium of the dynamics* (COSA)*. Then, the weighted sum of the differences between the signal and the state, $\sum_{i=1}^N d_i(s(x_i^*) - x_i^*)$, must be zero:*

$$\mathbf{1}^\top \boldsymbol{D}(\boldsymbol{s}(\boldsymbol{x}^*) - \boldsymbol{x}^*) = \sum_{i=1}^{N} d_i(s(x_i^*) - x_i^*) = 0.$$

Proof. Takes $\boldsymbol{w}$ be the left eigenvector of $\boldsymbol{D}^{-1}\boldsymbol{A}$ associated with the eigenvalue 1. Since the graph is connected and symmetric, one has $\boldsymbol{w} = \boldsymbol{D}\mathbf{1}$. Thus, the scalar product of $\boldsymbol{x}^*$ with $\boldsymbol{w}$ yields

$$\boldsymbol{w}^\top \boldsymbol{x}^* = \boldsymbol{w}^\top \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}^*) = \boldsymbol{w}^\top \boldsymbol{s}(\boldsymbol{x}^*).$$

Equivalently, one has $\boldsymbol{w}^\top(\boldsymbol{s}(\boldsymbol{x}^*) - \boldsymbol{x}^*) = 0 \Leftrightarrow \mathbf{1}^\top \boldsymbol{D}(\boldsymbol{s}(\boldsymbol{x^*}) - \boldsymbol{x}^*) = 0$. □

This lemma implies that at equilibrium, the total "pressure" exerted by agents on the system (quantified by $s(x_i^*) - x_i^*$ and weighted by degree) must sum to zero. While this provides a useful diagnostic, it does not guarantee that such equilibria exist. The following result, based on Brouwer's Fixed Point Theorem, establishes their existence.

Lemma 3.4.2 (Existence of Equilibria)**.** *The system* (COSA) *possesses at least one equilibrium point $\boldsymbol{x}^* \in \mathcal{X}$.*

Proof. Under Standing Assumption 3.2 on the signal function, s is a continuous function. Then, the map $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x})$ is continuous on the compact, convex set $\mathcal{X} = [-1, 1]^N$. Furthermore, as $\boldsymbol{D}^{-1}\boldsymbol{A}$ is a row-stochastic matrix and the range of s is contained within $[-1, 1]$ (i.e., $\boldsymbol{s}(\mathcal{X}) \subseteq \mathcal{X}$), the map $\boldsymbol{g}$ necessarily maps $\mathcal{X}$ to itself. i.e., $\boldsymbol{g}(\mathcal{X}) \subseteq \mathcal{X}$.

Then, by Brouwer's Fixed Point Theorem, the continuous map $\boldsymbol{g}$ must have at least one fixed point $\boldsymbol{x}^* \in \mathcal{X}$. This fixed point $\boldsymbol{x}^*$, satisfying $\boldsymbol{x}^* = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}^*)$, is precisely an equilibrium for the dynamics (COSA). □

Having guaranteed their existence, we can now classify equilibria based on the collective outcome they represent: perfect agreement versus persistent disagreement. This distinction is central to understanding the expressive power of the model.

Definition 3.4.3 (Classification of Equilibria: FSE and NFSE)**.** *Equilibria of the system* (COSA) *fall into one of two classes:*

- ***Fully Synchronized Equilibria (FSE):*** *States of complete agreement, $\boldsymbol{x}^* \in \mathcal{S}$, where all agents share a common value, $x_i^* = c$ for all $i \in \mathcal{V}$.*
- ***Non-Fully Synchronized Equilibria (NFSE):*** *States of persistent disagreement, $\boldsymbol{x}^* \in \mathcal{X} \backslash \mathcal{S}$, where agents' states settle to different values, often forming distinct clusters or patterns.*

This FSE/NFSE dichotomy forms the central analytical axis for the remainder of this thesis. It partitions the system's long-term behavior into two completely different regimes of either full synchronization or structured disagreement. To understand the conditions that lead to one regime or the other, we must analyze each class in detail. It is analytically prudent to begin with the simplest and most structured class of outcomes: the Fully Synchronized Equilibria.

3.5 The Fully Synchronized (FS) manifold and the FS dynamics

Our analysis now proceeds to the first and most central class of equilibria: the states of full synchronization (FSE). The study of these states is greatly simplified by recognizing that they all reside on a common, one-dimensional geometric structure within the state space.

We formally define this structure as the **Fully Synchronized Manifold (FSM)**. This section is dedicated to characterizing this manifold and, most importantly, the dynamics restricted to it. By reducing the N-dimensional problem to a simple scalar dynamic on this manifold, we can rigorously determine the stability and basins of attraction for all potential states of consensus.

Definition 3.5.1 (Fully Synchronized (FS) states and the Fully Synchronized Manifold (FSM))**.** *A state is a **fully synchronized (FS) state** (also called **consensus** state) if $x_i = x_j$ for all $i, j \in \mathcal{V}$.*

*The **fully synchronized manifold (FSM)** is defined as the set of all fully synchronized states:*

$$\mathcal{S} = \left\{ \boldsymbol{x} \in [-1,1]^N \mid \forall i,j \in \mathcal{V},\ x_i = x_j \right\}.$$

Notation 3.5.1. *For $c \in [-1,1]$, we denote by*

- *for a scalar $c \in [-1,1]$, $c\mathbf{1}$ the fully synchronized state at c,*
- *for an set $I \subset [-1,1]$, $\mathcal{S}[I] = \{\boldsymbol{x} \in \mathcal{S} \mid \exists y \in I,\ \boldsymbol{x} = y\mathbf{1}\}$ the fully synchronized manifold (FSM) intersecting I^N.*

For the analysis within this section, particularly concerning the dynamics *over* the Fully Synchronized Manifold $\mathcal{S}$, we can relax the strict requirements of Standing Assumption 3.1. The derivations here primarily rely on a more general condition for the communication topology.

Alternative Assumption 3.3 (No Isolated Agents)**.** *Every agent $i \in \mathcal{V}$ has at least one incoming connection, i.e., its in-degree $d_i = \sum_{j=1}^{N} a_{ji} > 0$.*

Alternative Assumption 3.3 ensures that the normalized adjacency matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$ is well-defined and row-stochastic, satisfying the property $\boldsymbol{D}^{-1}\boldsymbol{A}\mathbf{1} = \mathbf{1}$. This property is sufficient for analyzing the scalar dynamics of the common opinion on $\mathcal{S}$, as the specific graph structure (e.g., directedness, global connectivity) does not affect the self-consistent evolution of a shared state. It is crucial, however, to remember that Standing Assumption 3.1 remains in full effect for the analysis of the full multi-dimensional (COSA) system outside $\mathcal{S}$, especially when considering non-synchronized states and strong monotonicity properties (needed for the strong monotonicity property of the dynamics).

3.5.1 Invariance of the Fully Synchronized Manifold (FSM)

The most central property of the synchronized manifold is that it is a "trap" for the dynamics: once a trajectory enters this state of perfect agreement, it can never leave. Establishing this forward invariance is the key step that will later allow us to reduce the N-dimensional system (COSA) to a simple one-dimensional one for all synchronized states.

Proposition 3.5.2. *The synchronization manifold $\mathcal{S}$ is forward invariant for the dynamics* (COSA). *i.e., if $\boldsymbol{x}(0) \in \mathcal{S}$, then $\boldsymbol{x}(t) \in \mathcal{S}$ for all $t \geq 0$.*

Proof. To prove that the synchronization manifold $\mathcal{S}$ is forward invariant, we must show that for any trajectory $\boldsymbol{x}(t)$ starting in $\mathcal{S}$, it remains in $\mathcal{S}$ for all $t \geq 0$. This is equivalent to showing that if all agent states are identical at some time t, their velocities are also identical at that time, thus preserving their perfect agreement.

Let $\boldsymbol{x} \in \mathcal{S}$. Then, there exists a scalar $c \in [-1, 1]$ such that $\boldsymbol{x} = c\mathbf{1}$. This means $x_k = c$ for all agents $k \in \mathcal{V}$.

Let us examine the time derivative for an arbitrary agent $i \in \mathcal{V}$:

$$\begin{aligned}\dot{x}_i &= \frac{1}{d_i}\sum_{k=1}^{N} a_{ik}s(x_k) - x_i = \frac{1}{d_i}\sum_{k=1}^{N} a_{ik}s(c) - c \\ &= s(c)\left(\frac{1}{d_i}\sum_{k=1}^{N} a_{ik}\right) - c.\end{aligned}$$

As established in the Proposition 3.2.1-(ii), the matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$ is row-stochastic, which means each of its rows sums to one. The term in the parenthesis is precisely the sum of the i-th row of $\boldsymbol{D}^{-1}\boldsymbol{A}$. Therefore, for any agent i with $d_i > 0$ (as per Alternative Assumption 3.3), we have:

$$\frac{1}{d_i}\sum_{k=1}^{N} a_{ik} = 1.$$

Substituting this back into the expression for the velocity, we find:

$$\dot{x}_i(t) = s(c(t)) - c(t).$$

This result is crucial because the expression for the velocity $\dot{x}_i(t)$ is independent of the specific agent index i. It depends only on the common state $c(t)$. Therefore, $\dot{x}_i(t) = \dot{x}_j(t)$ for all $i, j \in \mathcal{V}$.

Since all agents starting on the manifold share the same initial state and have the same velocity at all times, their states remain identical. Thus, the synchronization manifold $\mathcal{S}$ is forward invariant. □

The invariance of the fully synchronized manifold is a cornerstone of the analysis. It identifies consensus as a structurally stable configuration of the system and, most importantly, allows us to study the behavior of FSM independently. Since trajectories cannot leave $\mathcal{S}$, the complex N-dimensional dynamics (COSA), when restricted to this manifold, must collapse into a much simpler form. The next section will exhibit this simplification.

3.5.2 The Fully Synchronized Dynamics

We now leverage this invariance to derive the precise form of these simplified dynamics. For any trajectory $\boldsymbol{x}(t) \in \mathcal{S}$, there exists a scalar trajectory $x(t) \in [-1, 1]$ such that $\boldsymbol{x}(t) = x(t)\mathbf{1}$. By substituting this into the main (COSA) equation, we can study the system onto a single scalar ordinary differential equation that governs the evolution of the common opinion. This reveals the core mechanisms of consensus in their simplest form (this will be the starting point of the analysis of the FSM in Chapter 4):

$$\begin{aligned}\dot{\boldsymbol{x}} &= \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x} = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(x\mathbf{1}) - x\mathbf{1} \\ &= \boldsymbol{D}^{-1}\boldsymbol{A}s(x)\mathbf{1} - x\mathbf{1} = s(x)(\boldsymbol{D}^{-1}\boldsymbol{A}\mathbf{1}) - x\mathbf{1} \\ &= s(x)\mathbf{1} - x\mathbf{1} = (s(x) - x)\,\mathbf{1}\end{aligned}$$

which is equivalent to the scalar dynamics:

$$\dot{x} = s(x) - x := f(x) \tag{FS}$$

This scalar dynamics governs the evolution of the common opinion x when all agents are synchronized.

Being able to restrict the analysis to the FSM simplifies the analysis of the asymptotic behavior on $\mathcal{S}$, guaranteeing that we only need to identify and analyze the stability of equilibrium points in (FS). It also highlights a key distinction: while the full multi-dimensional system (COSA) requires Lipschitz continuity for its general monotone flow property, the one-dimensional synchronized dynamics on $\mathcal{S}$ robustly exhibits monotonicity and lacks oscillations even with discontinuous, non-decreasing signal functions as we will see in the following sections.

Characterizing Equilibria on the FSM

We now analyze the equilibria of the one-dimensional dynamics (FS).

Proposition 3.5.3. *Let $s : [-1, 1] \to [-1, 1]$ be a signal function. Then, the equilibria of the dynamics (FS) are precisely the fixed points of s.*

Proof. An equilibrium $x^* \in [-1, 1]$ is defined by the condition $f(x^*) = 0$. From (FS), this is equivalent to $s(x^*) - x^* = 0$, which precisely means $s(x^*) = x^*$. Therefore, the set of equilibria is $\mathrm{Fix}(s)$. □

Remark 3.5.1. *This results is completely aligned with the classical linear consensus model where $s(x) = x$. In that particular case, any $x \in [-1, 1]$ is an equilibrium on $\mathcal{S}$, representing a continuum of possible consensus values. But for a general nonlinear $s(x)$, the set of equilibria is only restricted to the specific values where $s(x)$ intersects the identity line $y = x$. i.e., $x^* \in \mathrm{Fix}(s)$. This set is often the union of a finite number of points, meaning that the system can only converge to a finite number of consensus values.*

Since s is a non-decreasing function mapping the compact convex set $[-1,1]$ to itself, Lemma 3.2.4 guarantees that the set of fixed points $\mathrm{Fix}(s)$ is non-empty, and thus at least one equilibrium for (FS) is guaranteed to exist within $[-1,1]$.

Corollary 3.5.4. *Let $s : [-1,1] \to [-1,1]$ be a signal function. Then, the dynamics (FS) has at least one equilibrium.*

Proof. This is a direct consequence of Proposition 3.5.3 and Lemma 3.2.4. □

Invariant Intervals and Trajectory Localization

A key feature of the (FS) dynamics is that over $\mathcal{S}$, trajectories can be *trapped* within certain intervals. The boundaries of these intervals are determined by the points where the signal function $s(x)$ crosses the identity line (i.e., the fixed points of s).

Proposition 3.5.5 (Invariant Intervals)**.** *The following statements hold true:*

(i) For any $a, b \in [-1,1]$ with $a \leq b$, if $s(a) \geq a$ and $s(b) \leq b$, then the interval $[a,b]$ is forward invariant for the dynamics (FS).

(ii) In addition, if $s(a) > a$ and $s(b) < b$, then the open interval (a,b) is forward invariant for the dynamics (FS).

For the fixed points of s, this means that:

(i-bis) For any two fixed points $\underline{c}, \overline{c} \in \mathrm{Fix}(s)$ such that $\underline{c} \leq \overline{c}$, the interval $[\underline{c}, \overline{c}]$ is forward invariant for the dynamics (FS).

(ii-bis) In addition, if there exist $x_1, x_2 \in (\underline{c}, \overline{c})$ such that $s(x_1) > x_1$ and $s(x_2) < x_2$, then the open interval $(\underline{c}, \overline{c})$ is forward invariant for the dynamics (FS).

Proof. (i) The dynamics are governed by $\dot{x} = f(x) = s(x) - x$. For an interval $[a,b]$ to be forward invariant, by Nagumo's Theorem, the vector field must not point outward at its boundaries.

At the upper boundary $x = b$, the condition for invariance is $\dot{x}(b) \leq 0$. This is equivalent to $s(b) - b \leq 0$, or $s(b) \leq b$, which is given by the proposition's assumption.

At the lower boundary $x = a$, the condition for invariance is $\dot{x}(a) \geq 0$. This is equivalent to $s(a) - a \geq 0$, or $s(a) \geq a$, which is also given.

Since the vector field is non-positive at b and non-negative at a, any trajectory starting in $[a,b]$ cannot exit this interval.

(ii) Since s is non-decreasing, there exists a $\varepsilon > 0$ such that for all $y \in [b-\varepsilon, b]$, we have $s(y) < y$ and for all $y \in [a, a+\varepsilon]$, we have $s(y) > y$. Let us note $I_n = [a + \varepsilon/n, b - \varepsilon/n]$ for $n \geq 1$. For $n \geq 1$, the interval I_n is forward invariant and the sequence I_n is increasing. i.e., $I_n \subset I_{n+1}$. Since every I_n is forward invariant, so is their union:

$$\bigcup_{n \geq 1} I_n = \bigcup_{n \geq 1} [a + \frac{\varepsilon}{n}, b - \frac{\varepsilon}{n}] = (a,b).$$

Thus, the open interval (a,b) is forward invariant.

(i-bis)-(ii-bis) These items are a direct consequence of (i)-(ii). □

This result provides a powerful tool for localizing trajectories. It guarantees that the long-term behavior of a synchronized system is confined to the smallest such invariant interval containing its initial state. This naturally leads to the question of what happens inside these intervals.

Dynamical Behavior of the FS dynamics

The one-dimensional nature of the FS dynamics (FS) implies strong structural properties regarding its trajectories and *that even with discontinuous, non-decreasing signal functions*. Let us drop the Standing Assumption 3.2 for a moment and state the following proposition.

Proposition 3.5.6. *For the scalar dynamics (FS), given that s is non-decreasing on $[-1,1]$:*

(i) The dynamics cannot exhibit limit cycles or chaotic behavior.

(ii) Every trajectory $x(t)$ starting in $[-1,1]$ converges to a limit $x_\infty \in [-1,1]$.

(iii) Let $f(x) := s(x) - x$. A non-equilibrium trajectory (a trajectory not starting in an equilibrium point) is strictly monotonic.

- *If the trajectory is strictly increasing, its limit x_∞ satisfies $f(x_\infty^-) := \lim_{y \to x_\infty^-} f(y) = 0$.*
- *If the trajectory is strictly decreasing, its limit x_∞ satisfies $f(x_\infty^+) := \lim_{y \to x_\infty^+} f(y) = 0$.*

If additionally s is continuous, then

(ii-bis) Every trajectory starting in $[-1,1]$ converges to an equilibrium point in $\mathrm{Fix}(s)$.

(iii-bis) Trajectories exhibit monotonic behavior between equilibria: if $x_1^* < x_2^*$ are two consecutive equilibria, then for any $x_0 \in (x_1^*, x_2^*)$, the trajectory $x(t)$ will either monotonically increase towards x_2^* (if $f(x_0) > 0$) or monotonically decrease towards x_1^* (if $f(x_0) < 0$).

To simplify the proof, let us state and prove the following technical lemma.

Lemma 3.5.7. *Let $f(x) = s(x) - x$, where $s : [-1,1] \to [-1,1]$ is a non-decreasing function. If $f(a) > 0$ and $f(b) < 0$ for some $a < b$, then there exists a point $c \in (a,b)$ such that $f(c) = 0$.*

Proof. Let $C = \{x \in [a,b] \mid f(x) \geq 0\}$. The set C is non-empty since $a \in C$, and it is bounded above by b. By the least-upper-bound property of the real numbers, its supremum exists, let us define it as $c = \sup C$. By construction, $c \in [a,b]$. Since $f(a) > 0$ and $f(b) < 0$, we must have $c \in (a,b)$. Indeed, if we have $c = a$, we would have a contradition by taking the right-hand limit of $s(a^+) = \lim_{y \to a^+} s(y)$, which would be greater than $s(a^+) \geq s(a) > a$ (since $f(a) > 0$). This would yield $s(a^+) - a = f(a^+) > 0$, then there exists an $x \in (a, a^+)$ such that $f(x) > 0$, which contradicts the definition of c.

We now analyze the one-sided limits of $s(x)$ at c.

- For any $y \in (c,b]$, we have $y \notin C$, which implies $f(y) < 0$, or $s(y) < y$. Since s is non-decreasing, the right-hand limit $s(c^+) = \lim_{y \to c^+} s(y)$ exists. Taking the limit of the inequality gives $s(c^+) \leq c$.
- By the definition of the supremum, for any $\varepsilon > 0$, there exists an $x_\varepsilon \in C$ such that $c - \varepsilon < x_\varepsilon \leq c$. For this point, $f(x_\varepsilon) \geq 0$, so $s(x_\varepsilon) \geq x_\varepsilon$. Since s is non-decreasing, the left-hand limit $s(c^-) = \lim_{y \to c^-} s(y)$ exists and must satisfy $s(c^-) \geq c$.

For any non-decreasing function, we have the property $s(c^-) \leq s(c) \leq s(c^+)$. Combining our results yields the chain of inequalities:

$$c \leq s(c^-) \leq s(c) \leq s(c^+) \leq c.$$

This forces every term in the chain to be equal to c. In particular, $s(c) = c$, which means $f(c) = 0$. This proves that $f(x)$ cannot change sign on an interval without having a zero in that interval. □

We are now ready to prove the proposition.

Proof of Proposition 3.5.6. Let $f(x) := s(x) - x$.

(i) A trajectory $x(t)$ follows the vector field $f(x)$. For any non-equilibrium point $x(0)$, its velocity $f(x(0))$ is non-zero. By Lemma 3.5.7, the sign of $f(x)$ cannot change between any two consecutive equilibria. Thus, a trajectory starting at a non-equilibrium point is strictly monotonic. In a one-dimensional state space, trajectories cannot cross themselves, which excludes the existence of limit cycles or chaotic behavior.

(ii) If $x(0)$ is an equilibrium, the trajectory is constant and has converged. If not, the trajectory is strictly monotonic as established above. Since the state space $[-1, 1]$ is compact, the trajectory is bounded. By the Monotone Convergence Theorem, any bounded and monotonic function must converge to a finite limit. Therefore, $x_\infty = \lim_{t\to\infty} x(t)$ exists and lies in $[-1, 1]$.

(iii) Consider a strictly increasing trajectory where $\dot{x}(t) > 0$ and $\lim_{t\to\infty} x(t) = x_\infty$. As $t \to \infty$, $x(t)$ approaches x_∞ from the left. Since $s(x)$ is non-decreasing, the one-sided limits of $f(x)$ exist. The limit of the velocity is:

$$x_\infty = \lim_{t\to\infty} \dot{x}(t) = \lim_{t\to\infty} f(x(t)) = \lim_{y\to x_\infty^-} f(y) =: f(x_\infty^-).$$

Since $\dot{x}(t) > 0$, we must have $x_\infty \geq 0$. Assume for contradiction that $x_\infty > 0$. Then, there exists a time T such that $\dot{x}(t) > x_\infty/2$ for all $t \geq T$. Integrating from T to t gives $x(t) - x(T) > x_\infty/2(t - T)$. As $t \to \infty$, the right-hand side grows unboundedly, contradicting that $x(t)$ converges to a finite limit. Thus, we must have $x_\infty = 0$, which implies $f(x_\infty^-) = 0$. An analogous argument for a strictly decreasing trajectory shows its velocity must converge to $x_\infty = f(x_\infty^+) \leq 0$. The assumption $x_\infty < 0$ leads to a contradiction, so $x_\infty = 0$, which implies $f(x_\infty^+) = 0$.

(ii-bis) If s is continuous, then $f(x)$ is also continuous. For a continuous function, the one-sided limits equal the function value: $f(x_\infty^-) = f(x_\infty^+) = f(x_\infty)$. From (iii), the limit x_∞ of any trajectory must satisfy either $f(x_\infty^-) = 0$ or $f(x_\infty^+) = 0$. By continuity, both conditions imply $f(x_\infty) = 0$. Therefore, the limit x_∞ must be an equilibrium point in $\mathrm{Fix}(s)$.

(iii-bis) Let $x_1^* < x_2^*$ be two consecutive equilibria. Since $f(x)$ is continuous and has no zeros in (x_1^*, x_2^*), the Intermediate Value Theorem implies that $f(x)$ must have a constant sign on this interval. For any initial condition $x_0 \in (x_1^*, x_2^*)$, if $f(x_0) > 0$, the trajectory is strictly increasing and, by (ii-bis), must converge to the next available equilibrium, which is x_2^*. Conversely, if $f(x_0) < 0$, the trajectory is strictly decreasing and must converge to x_1^*. The dynamics are confined to $[-1, 1]$ as $\dot{x}(-1) = s(-1) - (-1) \geq 0$ and $\dot{x}(1) = s(1) - 1 \leq 0$. □

This powerful result simplifies the analysis of the asymptotic behavior on $\mathcal{S}$, guaranteeing that we only need to identify and analyze the stability of equilibrium points. It also highlights a key distinction: while the full multi-dimensional system (COSA) requires Lipschitz continuity for its general monotone flow property, the one-dimensional synchronized dynamics on $\mathcal{S}$ robustly exhibits monotonicity and lacks oscillations even with discontinuous, non-decreasing signal functions.

Remark 3.5.2. *It is important to note that the trajectories in the discontinuous case may not necessarily converge to an equilibrium point but can also converge to a limit point instead. This phenomenon is captured by the definition introduced by [105, 106, 124] of extended equilibria. There the authors prove the convergence to an extended equilibrium point, which is a limit point of the trajectory.*

Figure 3.2 provides a visual representation of $f(x) = s(x) - x$ and its associated vector field. Equilibria correspond to the points where the graph of $f(x) = s(x) - x$ intersects the x-axis. The direction of the vector field $\dot{x}$ (indicated by arrows) shows whether x increases or decreases. We

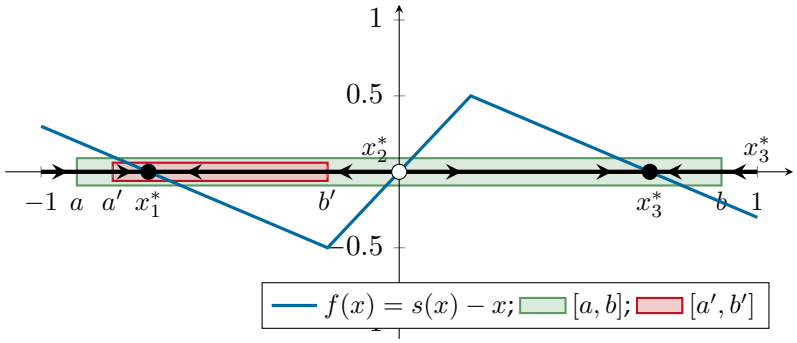


Figure 3.2: Illustration of the scalar dynamics $\dot{x} = s(x) - x$ for the signal function $s(x) = 0.7 \max(-1, \min(5x, 1))$. Equilibria are at the intersections (x_1^*, x_2^*, x_3^*). The vector field $f(x) = s(x) - x$ (indicated by arrows on the horizontal axis) shows the direction of movement. We present some invariant intervals of the dynamics to illustrate Proposition 3.5.5.

can then intuitively see that x_1^* and x_3^* are stable equilibria and x_2^* is an unstable equilibrium. In the next section, we will characterize the stability of these equilibria more rigorously.

3.5.3 Stability Analysis of Equilibria on the FS Manifold

The stability of an equilibrium x^* for a one-dimensional continuous-time dynamical system $\dot{x} = f(x)$ is determined by the sign of $f(x)$ in the immediate neighborhood of x^*. An equilibrium x^* is locally asymptotically stable if $f(x)$ points towards x^* from both sides, ensuring monotonic convergence. This condition is mathematically expressed as $(x - x^*)f(x) \leq 0$ for x in a sufficiently small neighborhood of x^*.

To characterize this behavior, we utilize the concepts of *stable estimation*, *underestimation*, and *overestimation* as introduced by Definition 3.2.2. The following theorem precisely characterizes the local stability properties of the equilibria of (FS) based on these signal function properties.

Theorem 3.5.8. *Let $x^* \in [-1, 1]$ be an equilibrium of the dynamics (FS). Then:*

*(i) x^* is **locally stable** if and only if x^* is a stable fixed point of s. i.e., $x^* \in \mathrm{Fix}^\bullet(s)$.*

*(ii) x^* is **locally asymptotically stable** if and only if x^* is a stable fixed point of s and x^* is isolated in $\mathrm{Fix}(s)$. i.e., $x^* \in \mathrm{Fix}^\bullet(s)$ and there exists a neighborhood U of x^* such that $U \cap \mathrm{Fix}(s) = \{x^*\}$.*

*(iii) x^* is **unstable** if and only if x^* is an unstable fixed point of s. i.e., $x^* \in \mathrm{Fix}^\circ(s)$.*

Proof. (i) By Definition 2.4.30, x^* is locally stable if and only if

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ such that } |x(0) - x^*| < \delta \implies |x(t) - x^*| < \varepsilon \text{ for all } t \geq 0.$$

This is equivalent to say that trajectories starting in a neighborhood of x^* remain in that neighborhood for all time. Since it is one dimensional, it means that for $x^* - \delta < x < x^*$, $f(x) \leq 0$ and for $x^* < x < x^* + \delta$, $f(x) \geq 0$. One has,

$$\begin{cases} f(x) \leq 0 \text{ for } x^* - \delta < x < x^* \\ f(x) \geq 0 \text{ for } x^* < x < x^* + \delta \end{cases} \Leftrightarrow \begin{cases} s(x) - x \leq 0 \text{ for } -\delta < x - x^* < 0 \\ s(x) - x \geq 0 \text{ for } 0 < x - x^* < \delta \end{cases}$$
$$\Leftrightarrow (x - x^*)(s(x) - x) \leq 0 \text{ for all } x \in [x^* - \delta, x^* + \delta].$$

Therefore, x^* is locally stable if and only if x^* is a stable fixed point of s. i.e., $x^* \in \mathrm{Fix}^\bullet(s)$.

(ii) For x^* to be locally asymptotically stable and attracting over a neighborhood U of x^*, let us consider the Lyapunov function

$$V(x) = \frac{1}{2}(x - x^*)^2 \geq 0.$$

The time derivative of V along the trajectories of the dynamics is given by

$$\dot{V}(x) = (x - x^*)(\dot{x} - \dot{x^*}) = (x - x^*)\dot{x} = (x - x^*)(s(x) - x).$$

Since x^* is an equilibrium, we have that $s(x^*) = x^*$. Therefore, V is a Lyapunov function for the dynamics (FS) and $\dot{V}(x^*) = 0$. To ensure x^* is locally asymptotically stable, we need to prove the existence of a neighborhood U of x^* such that $\dot{V}(x) < 0$ for all $x \in U \setminus \{x^*\}$. One has,

$$\dot{V}(x) < 0 \Leftrightarrow (x - x^*)(s(x) - x) < 0 \Leftrightarrow \mathrm{sgn}\,(x - x^*) \neq \mathrm{sgn}\,(s(x) - x) \text{ and } s(x) \neq x.$$

This means that x^* is locally asymptotically stable if and only if there exists a neighborhood U of x^* such that for all $x \in U \setminus \{x^*\}$, $\mathrm{sgn}\,(x - x^*) \neq \mathrm{sgn}\,(s(x) - x)$ and $s(x) \neq x$. The first condition means that x^* is a stable fixed point of s. i.e., $x^* \in \mathrm{Fix}^\bullet(s)$. The second condition means that x^* is isolated in $\mathrm{Fix}(s)$. i.e., there exists a neighborhood U of x^* such that $U \cap \mathrm{Fix}(s) = \{x^*\}$. Therefore, x^* is locally asymptotically stable if and only if x^* is a stable fixed point of s and x^* is isolated in $\mathrm{Fix}(s)$. i.e., $x^* \in \mathrm{Fix}^\bullet(s)$ and there exists a neighborhood U of x^* such that $U \cap \mathrm{Fix}(s) = \{x^*\}$.

(iii) Is given by the proof of (i) since (i) and (iii) are equivalent statements. Then, x^* is unstable if and only if x^* is an unstable fixed point of s. i.e., $x^* \in \mathrm{Fix}^\circ(s)$. □

Moreover, since the dynamics is 1-dimensional, we can use the Proposition 3.5.6 to characterize the attraction basin of equilibria.

Proposition 3.5.9 (Basin of Attraction). *Let s be a signal function and let $x^* \in \mathrm{Fix}(s)$ be an isolated fixed point of s. Let the basin of attraction of equilibrium x^* be denoted by $\mathcal{B}(x^*) = \{x_0 \in [-1, 1] \mid \lim_{t\to\infty} x(t) = x^*\}$.*

Define the boundaries for the basin as:

- $c_{\mathrm{L}}^* := \max(\{c \in \mathrm{Fix}^\circ(s) \mid x < x^*\} \cup \{-1\})$. *This is the largest unstable fixed point to the left of x^*, or -1 if none exists.*
- $c_{\mathrm{R}}^* := \min(\{c \in \mathrm{Fix}^\circ(s) \mid x > x^*\} \cup \{1\})$. *This is the smallest unstable fixed point to the right of x^*, or 1 if none exists.*

The following holds:

(i) $(c_{\mathrm{L}}^*, x^*] \subseteq \mathcal{B}(x^*)$ *if and only if for all* $x \in (c_{\mathrm{L}}^*, x^*)$, $s(x) > x$.

(ii) $[x^*, c_{\mathrm{R}}^*) \subseteq \mathcal{B}(x^*)$ *if and only if for all* $x \in (x^*, c_{\mathrm{R}}^*)$, $s(x) < x$.

(iii) $(x_L^*, x_R^*) = \mathcal{B}(x^*)$ *if and only if x^* is locally asymptotically stable.*

Proof. Let x^* be an isolated fixed point of s. This implies x^* is an equilibrium by Proposition 3.5.3. Moreover, since x^* is isolated, the quantities c_{L}^* and c_{R}^* are well-defined. Note that c_{L}^* and c_{R}^* are also equilibria of (FS) by Proposition 3.5.3. Thus, they cannot be in the attraction basin of x^*. i.e., $c_{\mathrm{L}}^* \notin \mathcal{B}(x^*)$ and $c_{\mathrm{R}}^* \notin \mathcal{B}(x^*)$. Moreover, by Proposition 3.5.5-(ii-bis), since $[c_{\mathrm{L}}^*, c_{\mathrm{R}}^*] \subseteq [-1, 1]$ is forward invariant and contains x^*, we have that $\mathcal{B}(x^*) \subseteq [c_{\mathrm{L}}^*, c_{\mathrm{R}}^*] \setminus \{c_{\mathrm{L}}^*, c_{\mathrm{R}}^*\} = (c_{\mathrm{L}}^*, c_{\mathrm{R}}^*)$.

Let us consider the function $V(x) = \frac{1}{2}(x - x^*)^2 \geq 0$. The time derivative of V along the trajectories of the dynamics is given by

$$\dot{V}(x) = (x - x^*)(s(x) - x).$$

Since x^* is an equilibrium, we have that $s(x^*) = x^*$. Therefore, $V(x)$ is a Lyapunov function for the dynamics (FS) and $\dot{V}(x^*) = 0$.

(i) ($\Rightarrow$) If $(c_{\mathrm{L}}^*, x^*] \subseteq \mathcal{B}(x^*)$, then for all $x(0) \in (c_{\mathrm{L}}^*, x^*]$, we have $\lim_{t\to\infty} x(t) = x^*$. Moreover, since for all $x \in (c_{\mathrm{L}}^*, x^*)$, we have $x < x^*$, we have that the function $t \mapsto x(t)$ must be increasing. Therefore, $\dot{x}(t) > 0$ for all $t \geq 0$. Then,

$$\dot{x} > 0 \Leftrightarrow s(x) - x > 0 \Leftrightarrow s(x) > x.$$

Therefore, $s(x) > x$ for all $x \in (c_{\mathrm{L}}^*, x^*)$.

($\Leftarrow$) Let $n_0 \in \mathbb{N}$ such that $[c_{\mathrm{L}}^* + 1/n_0, x^*] \subset (c_{\mathrm{L}}^*, x^*]$. Let us note $I_n = [c_{\mathrm{L}}^* + 1/n, x^*]$ for $n \geq n_0$. For $n \geq n_0$, the interval I_n is a compact forward invariant for the dynamics (FS).

Let us fix $n \geq n_0$. For $x \in I_n = [c_{\mathrm{L}}^* + 1/n, x^*]$, since $x \leq x^*$ we have

$$s(x) \geq x \Leftrightarrow (s(x) - x) \geq 0 \Leftrightarrow (x - x^*)(s(x) - x) \leq 0 \Leftrightarrow \dot{V}(x) \leq 0.$$

More specifically, we have $\dot{V}(x) = 0$ if and only if $x = x^*$. i.e., $\{x \in I_n \mid \dot{V}(x) = 0\} = \{x^*\}$. Therefore, by LaSalle's Invariance Principle, all trajectories starting in I_n converge to x^*. Thus, $I_n \subseteq \mathcal{B}(x^*)$.

Since this is true for all $n \geq n_0$, we have that every $I_n \subseteq \mathcal{B}(x^*)$ and thus

$$\bigcup_{n\geq n_0} I_n \subseteq \mathcal{B}(x^*) \Leftrightarrow \bigcup_{n\geq n_0} [c_{\mathrm{L}}^* + 1/n, x^*] \subseteq \mathcal{B}(x^*) \Leftrightarrow (c_{\mathrm{L}}^*, x^*] \subseteq \mathcal{B}(x^*).$$

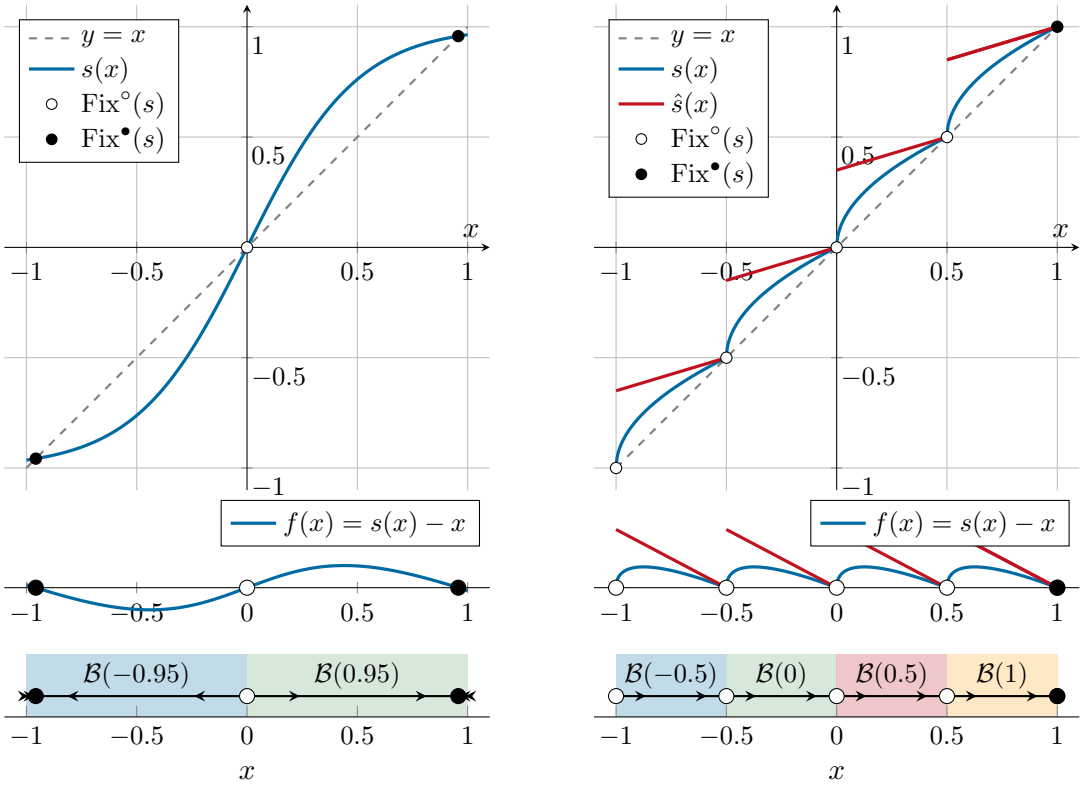
(ii) Similar argument to the proof of (i) applies by symmetry.

(iii) For x^* to be locally asymptotically stable, by proof of Theorem 3.5.8, it is equivalent to $x^* \in \mathrm{Fix}^\bullet(s)$ and x^* is isolated in $\mathrm{Fix}(s)$. This is exactly the conditions of (i) and (ii). Then, by (i) and (ii), we have that $(c_{\mathrm{L}}^*, c_{\mathrm{R}}^*) \subseteq \mathcal{B}(x^*)$. Since we already know that $\mathcal{B}(x^*) \subseteq (c_{\mathrm{L}}^*, c_{\mathrm{R}}^*)$, we have that $(c_{\mathrm{L}}^*, c_{\mathrm{R}}^*) = \mathcal{B}(x^*)$. Therefore, x^* is locally asymptotically stable if and only if $(c_{\mathrm{L}}^*, c_{\mathrm{R}}^*) = \mathcal{B}(x^*)$. □

The detailed analysis of the one-dimensional dynamics on the synchronization manifold $\mathcal{S}$ has provided key insights into the potential stability of the system's synchronized states. We established that the dynamics reduce to a simple scalar ordinary differential equation, $\dot{x}(t) = s(x(t)) - x(t)$, whose behavior is entirely determined by the fixed points of the signal function s. Crucially, we demonstrated that this system cannot exhibit complex behavior like limit cycles or chaotic trajectories, always converging to a limit point. The stability of these equilibria, *within the one-dimensional FS dynamics*, has been characterized and their basins of attraction are well-defined open intervals bounded by side unstable equilibria.

Figure 3.3 provides a visual synthesis of these theoretical results.

- Figure 3.3a shows a smooth, continuous signal function. In this scenario, we observe two stable equilibria (•) at the extremes, each attracting trajectories from a wide interval, and one unstable equilibrium (○) at the center acting as a repeller.
- Figure 3.3b illustrates a case with a signal function with succesive unstable fixed points. Here, multiple unstable equilibria (○) exist, yet they can still act as attractors from one side, creating distinct, contiguous basins of attraction. For instance, the interval $(0, 0.5)$ forms



(a) Case $s(x) = \tanh(2x)$.

(b) Case $s(x) = \sqrt{2x - \lfloor 2x \rfloor} + \lfloor 2x \rfloor$ and $\hat{s}(x) = \min(1, \lfloor 2x+1 \rfloor/2 + 0.3(x - \lfloor 2x+1 \rfloor/2))$

Figure 3.3: Illustration of the scalar dynamics $\dot{x} = s(x) - x$ for two different signal functions $s(x)$. From top to bottom, the first plot show the function $s(x)$ and the identity line $y = x$ and exhibits the fixed points of $s(x)$. The stable fixed points are represented by a black circle and the unstable ones by a white circle. The second plot show the dynamics function $f(x) = s(x) - x$ and the sign of the vector field $f(x)$. The third plot shows the basins of attraction of the different equilibria.

the basin of attraction for the unstable equilibrium at $x^* = 0.5$. The only asymptotically stable equilibrium at $x^* = 1$ collects trajectories from $(0.5, 1)$. Taking the example of the red line $\hat{s}(x)$ for a discontinuous signal function, the behavior is the same as for the continuous signal function: equilibria are the same and the basins of attraction are the same.

These examples highlight how the interplay between the signal function $s(x)$ and the identity line $y = x$ fully dictates the existence, stability, and regions of attraction for consensus states. Understanding these one-dimensional dynamics on $\mathcal{S}$ is a prerequisite for analyzing the more complex, multi-dimensional dynamics of the full system (COSA) and for investigating scenarios where agents might fail to reach full synchronization.

Remark 3.5.3 (The role of semi-stable equilibria)**.** *It is important to remark on the role of equilibria that are not asymptotically stable, as illustrated in Figure 3.3b. In this case, the equilibria at $x^* \in \{-1, -0.5, 0, 0.5\}$ are classified as unstable because they do not attract trajectories from both sides. For instance, any trajectory starting at $x_0 = 0.5 + \varepsilon$ for some small $\varepsilon > 0$ will move away from 0.5. According to Definition (iii), this is sufficient to classify the equilibrium as unstable.*

However, these unstable points can still possess a non-empty basin of attraction. As shown in the phase-line diagram for Case (b), the basin of attraction for the unstable equilibrium at $x^* = 0.5$ is the entire open interval $(0, 0.5)$. Trajectories starting in this interval monotonically increase and converge to 0.5. Equilibria that are attracting from one side and repelling from the other are often called ***semi-stable****.*

The proposition for the basin of attraction of an asymptotically stable *equilibrium remains correct because these semi-stable points serve as the boundaries. For the only asymptotically stable equilibrium at $x^* = 1$, its basin of attraction $\mathcal{B}(1)$ is correctly identified as the interval $(0.5, 1)$, where $x_L^* = 0.5$ is precisely the largest unstable (and in this case, semi-stable) equilibrium to its left.*

3.6 Preliminary Results on the COSA dynamics

In this part we use the insight gained from the one-dimensional dynamics (FS) on $\mathcal{S}$, to provide some preliminaries results for the full multi-dimensional (COSA) dynamics. Those results are used in the next Chapters to provide some results on the full multi-dimensional dynamics.

The main insight from the scalar system (FS) is the central role played by the fixed points of the signal function s. This section will demonstrate how these 1-dimensional properties form the very backbone of the N-dimensional system's structure.

First, in Subsection 3.6.1, we will formally establish that the FSE of the (COSA) system are a direct extension of the fixed points of s (in the sense of (FS)). Second, in Subsection 3.6.2, we will show that these same fixed points play a supplementary role: as they define the boundaries of invariant "trapping" regions (invariant hypercubes) for all trajectories of the system, as they did for the one-dimensional dynamics (FS).

3.6.1 Fully synchronized equilibria

We begin by characterizing the equilibria of the (COSA) system on the synchronization manifold $\mathcal{S}$. The following proposition provides the connection, demonstrating that the high-dimensional FSE are determined by the one-dimensional fixed points of the signal function s.

Proposition 3.6.1 (FSE are fixed points of s on $\mathcal{S}$)**.** *Under Alternative Assumption 3.3 (which implies $\boldsymbol{D}^{-1}\boldsymbol{A}\mathbf{1} = \mathbf{1}$), the set of FSE of the dynamics* (COSA) *is precisely the set defined by*

$$\mathcal{S}[\mathrm{Fix}(s)] = \{c\mathbf{1} \mid c \in \mathrm{Fix}(s)\}.$$

Proof. Let $\boldsymbol{x}^* \in \mathcal{S}$ be a synchronized equilibrium, meaning $\boldsymbol{x}^* = c\mathbf{1}$ for some common value $c \in [-1, 1]$. Substituting this into the equilibrium condition (3.9) on $\mathcal{S}$:

$$c\mathbf{1} = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(c\mathbf{1}).$$

Given that $\boldsymbol{s}(c\mathbf{1}) = s(c)\mathbf{1}$ (a vector where each component is $s(c)$) and, by Alternative Assumption 3.3, $\boldsymbol{D}^{-1}\boldsymbol{A}\mathbf{1} = \mathbf{1}$, the equation simplifies to:

$$c\mathbf{1} = s(c)\mathbf{1}.$$

This vector equality implies $c = s(c)$. Thus, c must be a fixed point of the signal function s, and any synchronized equilibrium $\boldsymbol{x}^*$ must be of the form $c\mathbf{1}$ where $c \in \mathrm{Fix}(s)$. Conversely, if $c \in \mathrm{Fix}(s)$, then $s(c) = c$. It is straightforward to verify that $\boldsymbol{x}^* = c\mathbf{1}$ satisfies the equilibrium condition $\boldsymbol{x}^* = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}^*)$, thereby being a synchronized equilibrium. □

Since Lemma 3.2.4 guarantees that at least one stable fixed point $c \in \mathrm{Fix}(s)$ always exists for a non-decreasing signal function, the existence of at least one FSE is immediately established.

Corollary 3.6.2. *The dynamics (COSA) always possesses at least one fully synchronized equilibrium. That is, there exists $c \in [-1, 1]$ such that $\boldsymbol{x}^* = c\mathbf{1}$ is an equilibrium.*

Proof. This is a direct consequence of Proposition 3.6.1 and Lemma 3.2.4. Lemma 3.2.4 guarantees that the set of fixed points $\mathrm{Fix}(s)$ is non-empty. For any $c \in \mathrm{Fix}(s)$, Proposition 3.6.1 confirms that $c\mathbf{1}$ is a synchronized equilibrium of the dynamics (COSA). □

This result emphasizes the role of the signal function's fixed points as the set of all possible consensus values for the network and point of main interest for the analysis of the (COSA) system.

3.6.2 Invariant Hypercubes Defined by Fixed Points of s

Now that we have established that FSEs correspond directly to the fixed points of s, we demonstrate their second role. These fixed points act as "walls" in the state space, extending the concept of invariant intervals from the 1-dimensional dynamics (FS) to invariant hypercubes in the full N-dimensional system (COSA). These regions provide a tool for predicting and localizing the system's long-term behavior.

Proposition 3.6.3. *The following statements hold true:*

(i) For any $a, b \in [-1, 1]$ with $a < b$, if $s(a) \geq a$ and $s(b) \leq b$, the hypercube $[a, b]^N$ is forward invariant under the dynamics (COSA).

(ii) For any $a, b \in [-1, 1]$ with $a < b$, if $s(a) > a$ and $s(b) < b$, the hypercube $(a, b)^N$ is forward invariant under the dynamics (COSA).

For the fixed points of s, this means that:

(i-bis) For any $\underline{c}, \overline{c} \in \mathrm{Fix}(s)$ such that $\underline{c} < \overline{c}$, the hypercube $[\underline{c}, \overline{c}]^N$ is forward invariant under the dynamics (COSA).

(ii-bis) In addition, if there exist $x_1 < x_2 \in (\underline{c}, \overline{c})$ such that $s(x_1) > x_1$ and $s(x_2) < x_2$, then the hypercube $(\underline{c}, \overline{c})^N$ is forward invariant under the dynamics (COSA).

Proof. The proof is a direct multi-dimensional extension of the argument used for Proposition 3.5.5.

(i) Since the vector field is Lipschitz continuous, by Nagumo's Theorem it is sufficient to analyze the dynamics at the boundary of the hypercube $[a, b]^N$.

Consider any agent $i \in \mathcal{V}$. Suppose its state starts at the upper boundary, $x_i = b$. The states of all other agents satisfy $x_j \leq b$. Since s is non-decreasing, $s(x_j) \leq s(b)$. The dynamics for agent i are:

$$\dot{x}_i = \frac{1}{d_i} \sum_{j=1}^{N} a_{ij} s(x_j) - b \leq \frac{1}{d_i} \sum_{j=1}^{N} a_{ij} s(b) - b$$

Since the matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$ is row-stochastic, $\sum_{j=1}^{N} a_{ij}/d_i = 1$. This simplifies the inequality to:

$$\dot{x}_i \leq s(b) - b$$

By assumption, $s(b) \leq b$, so $\dot{x}_i \leq 0$. This ensures the trajectory does not exit through the upper face $x_i = b$.

Symmetrically, suppose $x_i = a$. Then $x_j \geq a$ for all j, and thus $s(x_j) \geq s(a)$. The dynamics is:

$$\dot{x}_i = \frac{1}{d_i}\sum_{j=1}^{N} a_{ij} s(x_j) - a \geq \frac{1}{d_i}\sum_{j=1}^{N} a_{ij} s(a) - a = s(a) - a$$

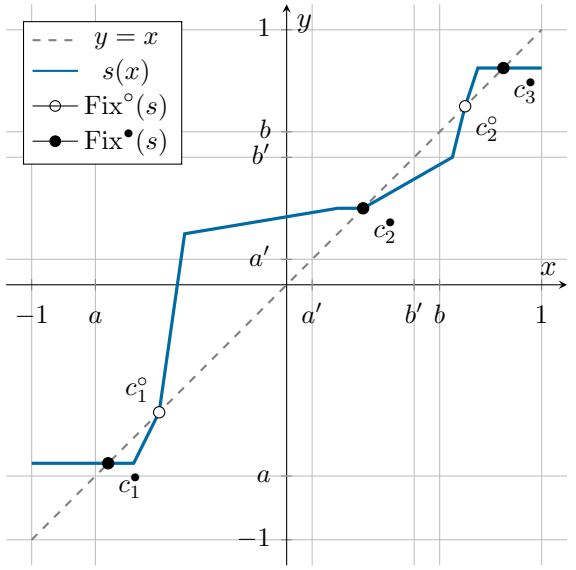
By assumption, $s(a) \geq a$, so $\dot{x}_i(t) \geq 0$. This ensures the trajectory does not exit through the lower face $x_i = a$. Since this holds for any agent i, the hypercube $[a, b]^N$ is forward invariant.

(ii) Since s is non-decreasing, there exists a $\varepsilon > 0$ such that for all $y \in [b - \varepsilon, b]$, we have $s(y) < y$ and for all $y \in [a, a + \varepsilon]$, we have $s(y) > y$. Let us note $H_n = [a + \varepsilon/n, b - \varepsilon/n]^N$ for $n \geq 1$. By (i), for $n \geq 1$, the hypercube H_n is forward invariant and the sequence H_n is increasing. i.e., $H_n \subset H_{n+1}$. Since every H_n is forward invariant, so is their union:

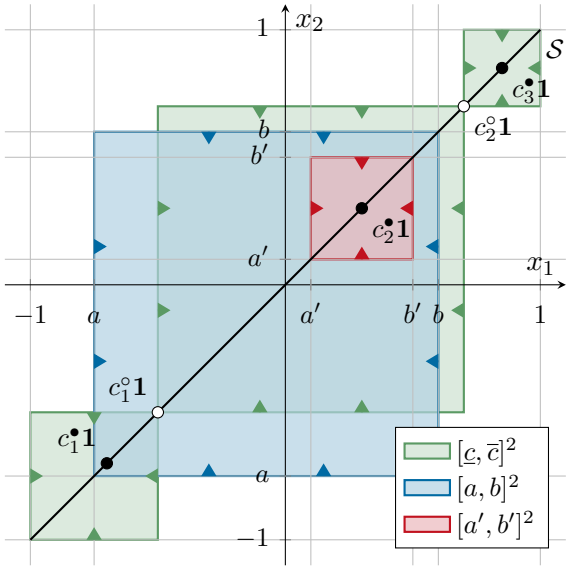
$$\bigcup_{n\geq 1} H_n = \bigcup_{n\geq 1} \left[a + \frac{\varepsilon}{n}, b - \frac{\varepsilon}{n}\right]^N = (a, b)^N.$$

Thus, the open hypercube $(a, b)^N$ is forward invariant.

(i-bis)-(ii-bis) These are direct corollaries, as any fixed point c satisfies $s(c) = c$. □



(a) A piecewise linear signal function $s(x)$ with two unstable fixed points c_1° and c_2° and three stable fixed points $c_1^{\bullet}$, $c_2^{\bullet}$, and $c_3^{\bullet}$. The shape of s hints the shape of the induced invariant hypercubes.



(b) The induced invariant hypercubes for the signal function of Figure 3.4a. The green area correspond to different invariant hypercubes defined by fixed points of s, while the blue area and the red area correspond to the invariant hypercube defined by the interval $[a, b]$ and $[a', b']$, respectively, illustrating Proposition 3.6.3(i) and (ii).

Figure 3.4: Illustration of Proposition 3.6.3 for a general signal function s. It is straightforward to observe the invariants of the dynamics (COSA) knowing the graph of the signal function s.

As an immediate consequence, every trajectory is trapped in any invariant hypercube defined by points a, b (or fixed points $\underline{c}, \overline{c}$) that contains its initial states.

Corollary 3.6.4. *Let $a, b \in [-1, 1]$ such that $a \leq \min_{i\in\mathcal{V}} x_i(0)$, $b \geq \max_{i\in\mathcal{V}} x_i(0)$, $s(a) \geq a$, and $s(b) \leq b$. Then, for all $t \geq 0$, $x_i(t) \in [a, b]$ for all $i \in \mathcal{V}$.*

In particular, if $\underline{c}, \overline{c} \in \mathrm{Fix}(s)$ such that $\underline{c} \leq \min_{i\in\mathcal{V}} x_i(0)$ and $\overline{c} \geq \max_{i\in\mathcal{V}} x_i(0)$, then for all $t \geq 0$, $x_i(t) \in [\underline{c}, \overline{c}]$ for all $i \in \mathcal{V}$.

The consequence of this proposition and its corollary is that every trajectory remains within the smallest hypercube (defined by two fixed points of s) that contains its initial states of all agents.

Our analysis in this section has thus revealed a clear picture of the system's global structure in (COSA). The state space is partitioned by a grid of invariant hypercubes, whose boundaries are defined by the system's own fully synchronized equilibria. This analysis yields the two central questions that will drive the subsequent chapters:

(i) **Synchronization:** Under what conditions are the FSEs stable attractors for trajectories within these hypercubes? This question of stability and global convergence to consensus will be the focus of the Chapter The Full Synchronization in COSA.

(ii) **Clustering:** What happens if the FSEs are not stable? Under what conditions do they lose stability? This opens the door to the existence of Non-Fully Synchronized Equilibria (NFSE), which we will explore next as a preview of the Chapter Partial Synchronization in COSA.

To motivate the next chapters, we will provide a constructive proof of the existence of an NFSE for the (COSA) system.

3.6.3 Non-Fully Synchronized Equilibria: A First Look

While global synchronization represents a state of perfect agreement, real-world collective behaviors often exhibit persistent disagreement, leading to the formation of distinct, locally synchronized clusters. These are the Non-Fully Synchronized Equilibria, where agents' states diverge into multiple values. The emergence and stability of NFSE are considerably more complex to characterize than FSE, as they arise from a interplay between network topology and the specific nonlinearities of the agent interactions (as we will see in Chapter 4). This section provides an initial exploration into NFSE, demonstrating their existence and highlighting the mechanisms that give rise to them, thereby laying the groundwork for a more comprehensive analysis in later chapters.

Example 3.6.1 serves as a foundational demonstration of how stable NFSE can emerge in a networked system. It reveals that the loss of stability of an FSE can catalyze the birth of these more complex, clustered states. This phenomenon is linked to specific properties of the signal function and the network structure, as illustrated by a local bifurcation analysis in Figure 3.5.

Example 3.6.1 (Emergence of a Stable NFSE on a Line Graph)**.** *To provide a constructive proof that* stable *NFSE can exist, we consider a line graph $\mathcal{G}$ with $N = 5$ agents and an odd signal function, $s(-x) = -s(x)$. This system possesses an anti-symmetric invariant subspace,*

$$\mathcal{M} = \{\boldsymbol{x} \in \mathcal{X} \mid x_3 = 0, x_2 = -x_4, \text{ and } x_1 = -x_5\}.$$

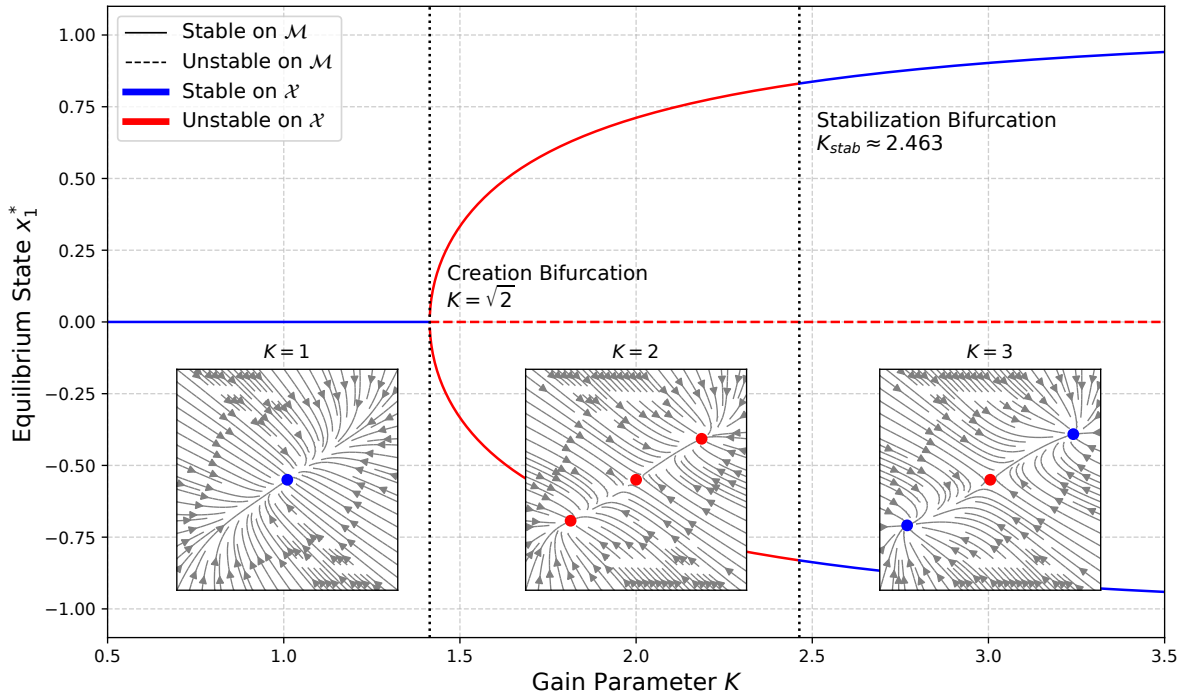


Figure 3.5: Bifurcation diagram for the $N = 5$ line graph, illustrating the distinction between stability on the invariant manifold $\mathcal{M}$ and stability in the full state space $\mathcal{X}$. The plot shows the equilibrium state x_1^* as a function of the gain parameter K. *Stability on $\mathcal{M}$ (lines):* The FSE at $x_1^* = 0$ loses stability at $K = \sqrt{2}$ (dashed line), creating two branches of NFSE that are stable on the manifold (solid lines). *Stability in $\mathcal{X}$ (red/blue colors):* The FSE becomes unstable in the full space at $K = \sqrt{2}$ (red dashed line). Crucially, the new NFSE branches are also born unstable in $\mathcal{X}$ (solid red lines), despite their stability on $\mathcal{M}$. They only become stable in the full space (solid blue lines) after the transversal stabilization bifurcation at $K_{\text{stab}} \approx 2.463$. The inset phase portraits visualize the dynamics (3.11) on $\mathcal{M}$ at three representative values of K.

Then, the set $\mathcal{M}$ is an invariant manifold under the dynamics (COSA). *Indeed, for any $\boldsymbol{x} \in \mathcal{M}$:*

$$\begin{aligned}
\dot{x}_1 &= s(x_2) - x_1 = -s(-x_2) + x_5 = -s(x_4) + x_5 = -\dot{x}_5 \\
\dot{x}_2 &= \frac{s(x_1) + s(x_3)}{2} - x_2 = -\left(\frac{s(-x_1)}{2} + x_4\right) = -\dot{x}_4 \\
\dot{x}_3 &= \frac{s(x_2) + s(x_4)}{2} - x_3 = \frac{s(x_2) + s(-x_2)}{2} - 0 = 0
\end{aligned}$$

Therefore, any trajectory starting on $\mathcal{M}$ remains on $\mathcal{M}$. The dynamics on this manifold reduce to the two-dimensional system in variables $(\tilde{x}_1, \tilde{x}_2) = (x_1, x_2)$ given by

$$\begin{cases} \dot{\tilde{x}}_1 = s(\tilde{x}_2) - \tilde{x}_1 \\ \dot{\tilde{x}}_2 = \frac{1}{2} s(\tilde{x}_1) - \tilde{x}_2 \end{cases} \tag{3.11}$$

The stability of equilibria in this system depends critically on the gain at the origin, $K = s'(0)$. A local bifurcation analysis, visualized comprehensively in Figure 3.5, reveals a two-stage process. First, at $K_{\text{bif}} = \sqrt{2}$, the FSE becomes unstable on $\mathcal{M}$, giving rise to two symmetric branches of NFSE. However,

these equilibria remain unstable in the full state space $\mathcal{X}$ due to dynamics transverse to the manifold. Second, at a higher critical gain K_{stab}, a transversal stabilization bifurcation occurs, rendering the NFSE locally stable in $\mathcal{X}$.

Example 3.6.1 is not an isolated curiosity. It is rather a basic example of how stable NFSE can emerge in a networked system. First, the instability of an FSE (e.g., at the origin) is a prerequisite for the system to "break" full synchronization and explore other equilibrium states. Second, the emergent NFSE are typically characterized by a structured partitioning of agents' states, forming distinct clusters. Third, the specific shape and parameters of the nonlinear signal function, particularly its local gain (slope), dictate the transition from synchronization to clustering. These observations collectively indicate that NFSE are not merely states of persistent disagreement, but often represent robust, structured patterns of "agreed-upon dissensus" within the network. Understanding the conditions for their emergence, their stability, and how to quantify their robustness against external perturbations will be the central focus of the subsequent chapters.

3.7 Concluding Remarks and Outlook

This chapter has laid the analytical foundations for the Continuous Opinion Signaled Action (COSA) model, establishing it as a rigorous and versatile framework for studying multi-agent systems with nonlinear interactions. By generalizing discrete-action protocols to a broad class of non-decreasing, Lipschitz-continuous signal functions, the (COSA) model provides a unified basis for analyzing collective behaviors from full consensus to structured disagreement.

The analysis was built upon a cornerstone property: the dynamics are inherently **cooperative** (Proposition 3.3.2). This structural feature is as the core of the entire framework, as it guarantees that the system's flow is monotone and precludes complex non-convergent behaviors such as limit cycles or chaos. This powerful result ensures that all trajectories eventually converge to a well-defined set of equilibria, shifting the analytical focus from the question of *if* the system settles to the more crucial question of *where* it settles.

The key to unlocking the system's behavior was the analysis of its dynamics on the one-dimensional **Fully Synchronized (FS) Manifold** with the dynamics (FS). This reduction revealed the central insight of the chapter: the **dual role of the signal function's fixed points**. First, we established that the set of all fully synchronized equilibria (FSE) is precisely determined by the fixed points of s (Proposition 3.6.1). Second, and more critically, we demonstrated that these fixed points impose a global structure on the full N-dimensional state space. They serve as the boundaries for a grid of **invariant hypercubes** that partition the space and permanently trap all system trajectories (Proposition 3.6.3).

Thus, the high-dimensional collective dynamics of the (COSA) model are shown to be governed by the simple, scalar properties of the individual agent's signal function. This understanding, encapsulated by the FSE/NFSE dichotomy (Definition 3.4.3), provides the analytical toolkit for the remainder of this thesis.

With this framework established, the subsequent chapters will explore the two central regimes of collective behavior: Full Synchronization (FS) and Partial Synchronization (PS) in (COSA).

Chapter 4

The Full Synchronization in COSA

4.1 Introduction

The preceding chapter established the foundational properties of the (COSA) model, concluding that its cooperative nature ensures the convergence of trajectories to a well-defined set of equilibria. This finding partitions the system's long-term behavior into two regimes: full synchronization, where all agents reach a FSE, and structured disagreement, where they settle into persistent, clustered states NFSE. This chapter undertakes a systematic and comprehensive analysis of the first regime, addressing the question: under what conditions is full synchronization the unique and mandatory outcome of the network dynamics?

Our investigation proceeds in a structured manner, building from local to global properties. We begin by establishing a rigorous framework for quantifying disagreement. With this analytical tool at hand, we first address the question of local stability. We will prove a powerful reduction principle: the stability of any FSE in the full N-dimensional space is completely determined by the stability of its corresponding fixed point in the fully synchronized manifold (FS). Having characterized the local behavior, we then turn to the chapter's central objective: to derive verifiable, network-wide conditions for global convergence to the FSM. This search leads us to the derivation of a sharp threshold that balances the cohesive strength of the network topology against the divisive potential of the signal function's nonlinearity.

Finally, we strengthen our analysis by examining the rate of convergence to the FSM, establishing a stricter condition that guarantees not just global attractivity but the global exponential stability of the synchronized manifold. By the conclusion of this chapter, we will have provided a complete analytical account of the conditions under which the COSA model achieves full and robust synchronization, thereby laying the essential groundwork for the subsequent analysis of disagreement.

4.2 Measuring Disagreement: The Natural Geometry of COSA

To rigorously analyze the system's convergence to the fully synchronized manifold $\mathcal{S}$, we must first establish a formal measure of the disagreement among agents. This measure is captured by the "distance" of a state vector $\boldsymbol{x} \in \mathcal{X}$ from the manifold $\mathcal{S}$. The choice of geometry, as defined by the choice of inner product, is not just a technical detail but is important to construct a clear and analytically tractable stability analysis of (COSA) around the FSM.

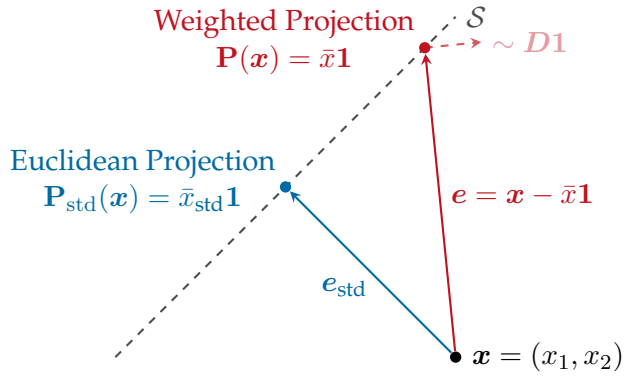


Figure 4.1: Geometric interpretation of projections onto the fully synchronized manifold $\mathcal{S}$ for $N = 2$ agents, assuming $d_1 > d_2$. The standard (Euclidean) projection $P(\boldsymbol{x})$ results in an error $\boldsymbol{e}_{\text{std}}$ that is visually perpendicular to $\mathcal{S}$. The weighted ($\boldsymbol{D}$-norm) projection $P_{\boldsymbol{D}}(\boldsymbol{x})$ results in the error $\boldsymbol{e}$ used in our analysis. This error is *not* Euclidean-perpendicular to $\mathcal{S}$, but it is $\boldsymbol{D}$-orthogonal to it (i.e., $\langle \boldsymbol{e}, \mathbf{1} \rangle_{\boldsymbol{D}} = e_1 d_1 + e_2 d_2 = 0$).

4.2.1 Motivation for a Weighted Inner Product

The most naive approach is to consider the system in the standard Euclidean space, equipped with the inner product $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^\top \boldsymbol{v}$. This perspective defines the standard average opinion

$$\bar{x}_{\text{std}} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

and a corresponding projection onto $\mathcal{S}$, $\boldsymbol{P}_{\text{std}}(\boldsymbol{x}) = \bar{x}_{\text{std}}\mathbf{1}$. The resulting error vector is orthogonal to the manifold $\mathcal{S}$ in the familiar geometric sense.

However, a closer look at the (COSA) dynamics reveals a misalignment to the Euclidean geometry. The key operator, the row-stochastic matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$, is generally *not* symmetric. Consequently, it is not self-adjoint with respect to the standard Euclidean inner product. This lack of symmetry complicates stability analysis, particularly within a Lyapunov framework where operator properties are central.

This analytical challenge motivates a search for a geometry in which the system's structure is naturally aligned with the dynamics. The most convenient choice is a **weighted inner product** induced by the degree matrix $\boldsymbol{D}$:

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\boldsymbol{D}} = \boldsymbol{u}^\top \boldsymbol{D} \boldsymbol{v}. \tag{4.1}$$

Under Standing Assumption 3.1 (undirected graph), the operator $\boldsymbol{D}^{-1}\boldsymbol{A}$ becomes self-adjoint with respect to this inner product (Proposition 3.2.1). This makes $\langle \cdot, \cdot \rangle_{\boldsymbol{D}}$ the **natural geometry** for the (COSA), as it aligns with the properties of its dynamic operators. Both projections are illustrated in Figure 4.1.

4.2.2 Weighted Projections and Equivalence of Convergence

Operating within this weighted geometry, we can now define a projection onto the fully synchronized manifold $\mathcal{S}$ that respects the system's inherent structure. The projection of a state $\boldsymbol{x}$ is the point on $\mathcal{S}$ closest to $\boldsymbol{x}$ as measured by the $\boldsymbol{D}$-norm, which leads to the following definitions.

Definition 4.2.1 (Weighted Average and Projection)**.** *The **weighted average** of the state $\boldsymbol{x}$ is defined as:*

$$\bar{x} = \frac{\mathbf{1}^\top \boldsymbol{D}\boldsymbol{x}}{\mathbf{1}^\top \boldsymbol{D}\mathbf{1}} = \frac{\sum_{i=1}^N d_i x_i}{\sum_{i=1}^N d_i}. \tag{4.2}$$

*This defines the **$\boldsymbol{D}$-orthogonal projection** of $\boldsymbol{x}$ onto $\mathcal{S}$:*

$$\mathbf{P}(\boldsymbol{x}) = \bar{x}\mathbf{1} \in \mathcal{S}.$$

This average is well-defined under Standing Assumption 3.1, which ensures $d_i > 0$ for all $i \in \mathcal{V}$.

While we adopt this weighted framework for its analytical advantages, it is required to establish that this choice does not alter the concept of convergence to the FSM. The following proposition confirms that convergence to synchronization in our chosen weighted geometry is perfectly equivalent to convergence in the Euclidean sense.

Proposition 4.2.2 (Topological Equivalence of Synchronization Distances)**.** *Let $d_{\text{std}}(\boldsymbol{x}, \mathcal{S}) = \|\boldsymbol{x} - \mathbf{P}_{\text{std}}(\boldsymbol{x})\|_2$ be the standard Euclidean distance to $\mathcal{S}$, and let $d_{\boldsymbol{D}}(\boldsymbol{x}, \mathcal{S}) = \|\boldsymbol{x} - \mathbf{P}(\boldsymbol{x})\|_{\boldsymbol{D}}$ be the weighted distance to $\mathcal{S}$. For any trajectory $\{\boldsymbol{x}(t)\}_{t\geq 0}$:*

$$\lim_{t\to\infty} d_{\text{std}}(\boldsymbol{x}(t), \mathcal{S}) = 0 \quad \iff \quad \lim_{t\to\infty} d_{\boldsymbol{D}}(\boldsymbol{x}(t), \mathcal{S}) = 0.$$

Proof. In the finite-dimensional space $\mathbb{R}^N$, all vector norms are equivalent. Since the degree matrix $\boldsymbol{D}$ is symmetric and positive definite (as $d_i > 0$ for a connected graph), the $\boldsymbol{D}$-norm, $\|\boldsymbol{v}\|_{\boldsymbol{D}} = \sqrt{\boldsymbol{v}^\top \boldsymbol{D}\boldsymbol{v}}$, is a valid norm. Thus, there exist constants $c_1, c_2 > 0$ such that $c_1\|\boldsymbol{v}\|_2 \leq \|\boldsymbol{v}\|_{\boldsymbol{D}} \leq c_2\|\boldsymbol{v}\|_2$ for all $\boldsymbol{v} \in \mathbb{R}^N$.

A trajectory converges to the manifold $\mathcal{S}$ if and only if the length of its component orthogonal to $\mathcal{S}$ converges to zero. Since both $d_{\text{std}}(\boldsymbol{x}, \mathcal{S})$ and $d_{\boldsymbol{D}}(\boldsymbol{x}, \mathcal{S})$ are zero if and only if $\boldsymbol{x} \in \mathcal{S}$, and because the norms are equivalent, convergence to zero in one metric implies convergence to zero in the other. Both distances therefore define the same topology on the state space and an identical notion of convergence to the subspace $\mathcal{S}$. □

Geometrically, the distance between $\mathbf{P}_{\text{std}}(\boldsymbol{x})$ and $\mathbf{P}(\boldsymbol{x})$ when $\boldsymbol{x}$ is on decreasing the closer the trajectory gets to the FSM, as it can be intuited from Figure 4.1.

This proposition provides a important guarantee: by adopting the dynamically motivated weighted projection, we gain analytical tractability without losing connection to the Euclidean understanding of convergence to the FSM.

4.2.3 The Synchronization Error Vector

With the weighted geometry fully motivated and validated, we can now formally define the central object for our stability analysis: the synchronization error vector $\boldsymbol{e}(t)$.

Definition 4.2.3 (Synchronization Error Vector)**.** *The **synchronization error vector**, $\boldsymbol{e}$, is the component of the state vector $\boldsymbol{x}$ that is $\boldsymbol{D}$-orthogonal to the fully synchronized manifold $\mathcal{S}$:*

$$\boldsymbol{e} = \boldsymbol{x} - \mathbf{P}(\boldsymbol{x}) = \boldsymbol{x} - \bar{x}\mathbf{1}. \tag{4.3}$$

By construction, this error vector has the property of being $\boldsymbol{D}$-orthogonal to any vector in $\mathcal{S}$, most notably the vector $\mathbf{1}$. This property is key for the stability analysis of the FSM as we will see later in this chapter.

Lemma 4.2.4 (Error Orthogonality)**.** *The synchronization error vector $\boldsymbol{e}$ is $\boldsymbol{D}$-orthogonal to the fully synchronized manifold $\mathcal{S}$. This is equivalent to the condition:*

$$\langle \boldsymbol{e}, \mathbf{1} \rangle_{\boldsymbol{D}} = \mathbf{1}^\top \boldsymbol{D} \boldsymbol{e} = \sum_{i=1}^{N} d_i e_i = 0.$$

Proof. From the definition of $\boldsymbol{e}$ and $\bar{x}$:

$$\mathbf{1}^\top \boldsymbol{D} \boldsymbol{e} = \mathbf{1}^\top \boldsymbol{D}(\boldsymbol{x} - \bar{x}\mathbf{1}) = \mathbf{1}^\top \boldsymbol{D} \boldsymbol{x} - \bar{x}(\mathbf{1}^\top \boldsymbol{D} \mathbf{1}).$$

Substituting the definition of the weighted average $\bar{x}$ from (4.2) yields:

$$\mathbf{1}^\top \boldsymbol{D} \boldsymbol{e} = \mathbf{1}^\top \boldsymbol{D} \boldsymbol{x} - \left(\frac{\mathbf{1}^\top \boldsymbol{D} \boldsymbol{x}}{\mathbf{1}^\top \boldsymbol{D} \mathbf{1}} \right) (\mathbf{1}^\top \boldsymbol{D} \mathbf{1}) = 0.$$

The result follows immediately. □

This orthogonal decomposition is the base of our analysis. It allows the state vector $\boldsymbol{x}$ to be cleanly separated into two components: its projection onto the fully synchronized manifold, $\bar{x}\mathbf{1}$, and its disagreement component, $\boldsymbol{e}$. meaning that $\boldsymbol{x}$ can be written as $\boldsymbol{x} = \mathbf{P}(\boldsymbol{x}) + \boldsymbol{e}$.

The squared magnitude of this disagreement component, measured in the weighted geometry, forms an ideal candidate for a Lyapunov function to measure the distance to the FSM:

$$V(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \bar{x}\mathbf{1})^\top \boldsymbol{D}(\boldsymbol{x} - \bar{x}\mathbf{1}) = \frac{1}{2}\boldsymbol{e}^\top \boldsymbol{D} \boldsymbol{e} = \frac{1}{2}\|\boldsymbol{e}\|_{\boldsymbol{D}}^2. \tag{4.4}$$

Since $\boldsymbol{D}$ is positive definite, $V(\boldsymbol{x}) \geq 0$ with $V(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x} = \bar{x}\mathbf{1}$, which corresponds precisely to the fully synchronized state where $\boldsymbol{x} \in \mathcal{S}$. This function, which measures the degree-weighted variance of opinions, will be the primary tool for characterizing the stability of synchronized equilibria in the next section. Additionally, we can define the distance to a particular FSE $\boldsymbol{x}^* = c\mathbf{1}$ with the natural weighted geometry:

$$d(\boldsymbol{x}, \boldsymbol{x}^*) = \|\boldsymbol{x} - \boldsymbol{x}^*\|_{\boldsymbol{D}} = \left((\boldsymbol{x} - \boldsymbol{x}^*)^\top \boldsymbol{D} (\boldsymbol{x} - \boldsymbol{x}^*) \right)^{1/2}.$$

which induce the following Lyapunov function:

$$V_c(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{D}(\boldsymbol{x} - c\mathbf{1}) = \frac{1}{2}\|\boldsymbol{x} - c\mathbf{1}\|_{\boldsymbol{D}}^2. \tag{4.5}$$

4.3 Stability and Attraction of Fully Synchronized Equilibria (FSE)

Having established a framework for measuring disagreement, we now turn to the question of stability. In Chapter 3, we fully characterized the stability and attraction basin of equilibria for the one-dimensional dynamics restricted to the fully synchronized manifold $\mathcal{S}$. The central result of this section is to demonstrate that this one-dimensional intuition extends to the full N-dimensional state space: the stability of a synchronized equilibrium $\boldsymbol{x}^* = c\mathbf{1}$ is determined entirely by the local properties of the signal function s at the fixed point c. Moreover, the attraction basin of the synchronized equilibrium $\boldsymbol{x}^* = c\mathbf{1}$ is determined by the local properties of neighboring fixed points of s as in the dynamics (FS).

4.3.1 Local Stability of FSE

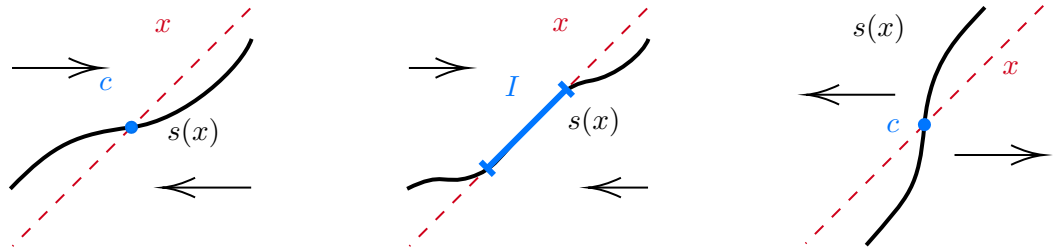
The following theorem provides a complete characterization of local stability, directly linking the (COSA) behavior to the classification of fixed points established in Definition 3.2.3.

Theorem 4.3.1. *Let $\boldsymbol{x}^* = c\mathbf{1}$ be a FSE. Then the following holds true:*

(i) $\boldsymbol{x}^*$ *is* ***locally stable*** *if and only if c is a stable fixed point of s, i.e., $c \in \mathrm{Fix}^\bullet(s)$.*

(ii) $\boldsymbol{x}^*$ *is* ***locally asymptotically stable*** *if and only if c is a stable fixed point of s and c is an isolated in $\mathrm{Fix}^\bullet(s)$.*

(iii) $\boldsymbol{x}^*$ *is* ***unstable*** *if and only if c is an unstable fixed point of s, i.e., $c \in \mathrm{Fix}^\circ(s)$.*



(a) Asymptotically stable. (b) Stable. (c) Unstable

Figure 4.2: Illustration of the local stability conditions for synchronization equilibria as described in Theorem 4.3.1. The red dashed line represents the identity function $y = x$, and the blue curve represents the signal function $s(x)$. Fixed points (intersections) are marked with blue circles. Arrows on the x-axis indicate the direction of change in the scalar dynamics $\dot{x} = s(x) - x$.

Proof. The proof connects the stability of the N-dimensional system near $\mathcal{S}$ to the stability of the one-dimensional FS dynamics (FS), $\dot{x} = s(x) - x$, which was fully characterized in Section 3.5.

(i) ($\Rightarrow$) Assume $\boldsymbol{x}^* = c\mathbf{1}$ is locally stable. This implies that trajectories starting in a neighborhood of $\boldsymbol{x}^*$ remain close. In particular, this must hold for trajectories that start on the FSM $\mathcal{S}$ itself. The dynamics on $\mathcal{S}$ is governed by $\dot{x} = s(x) - x$. By the stability conditions for one-dimensional systems (derived from Theorem 3.5.6), the local stability of c for this scalar system requires that in a neighborhood of c, we have $(x - c)(s(x) - x) \leq 0$. This is precisely the definition of c being a stable fixed point, $c \in \mathrm{Fix}^\bullet(s)$.

($\Leftarrow$) Assume $c \in \mathrm{Fix}^\bullet(s)$. This means there exists a neighborhood I of c where $(x - c)(s(x) - x) \leq 0$.

Consider the candidate Lyapunov function defined in (4.5) for analyzing the stability of the synchronized state $\boldsymbol{x}^* = c\mathbf{1} \in \mathcal{S}$, given by:

$$V_c(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{D}(\boldsymbol{x} - c\mathbf{1}).$$

This function measures the squared $\boldsymbol{D}$-weighted distance from the current state $\boldsymbol{x}$ to $c\mathbf{1}$.

The time derivative of $V_c(\boldsymbol{x})$ along the trajectories of dynamics (COSA) is given by:

$$\dot{V}_c(\boldsymbol{x}) = (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{D}\dot{\boldsymbol{x}} = (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{D}(\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}) = (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{D}\boldsymbol{x}.$$

Using the definition of the Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, we can rearrange the terms:

$$\begin{aligned}\dot{V}_c(\boldsymbol{x}) &= (\boldsymbol{x} - c\mathbf{1})^\top (\boldsymbol{D} - \boldsymbol{L})\boldsymbol{s}(\boldsymbol{x}) - (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{D}\boldsymbol{x} \\ &= (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{D}\boldsymbol{s}(\boldsymbol{x}) - (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}) - (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{D}\boldsymbol{x} \\ &= (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{D}(\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}) - (\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}).\end{aligned}$$

Since the graph $\mathcal{G}$ is undirected (Standing Assumption 3.1), the Laplacian matrix $\boldsymbol{L}$ is symmetric. Also, $\mathbf{1}$ is in the null space of $\boldsymbol{L}$ ($\boldsymbol{L}\mathbf{1} = \mathbf{0}$), implying $\mathbf{1}^\top \boldsymbol{L} = \mathbf{0}^\top$. This allows us to simplify the second term:

$$(\boldsymbol{x} - c\mathbf{1})^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}) - c\mathbf{1}^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}).$$

Recalling the quadratic form of $\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x})$ for symmetric $\boldsymbol{A}$ and non-decreasing s:

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}) = \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x_i - x_j)(s(x_i) - s(x_j)).$$

Therefore, the time derivative of $V_c(\boldsymbol{x})$ is:

$$\dot{V}_c(\boldsymbol{x}) = \underbrace{\sum_{i=1}^{N} d_i(x_i - c)(s(x_i) - x_i)}_{V_{c,1}(\boldsymbol{x})} \underbrace{- \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x_i - x_j)(s(x_i) - s(x_j))}_{V_{c,2}(\boldsymbol{x})}.$$

For $\boldsymbol{x}$ in a sufficiently small neighborhood of $c\mathbf{1}$, all x_i will be in the neighborhood I of c.

The first term,

$$V_{c,1}(\boldsymbol{x}) = \sum_{i=1}^{N} d_i(x_i - c)(s(x_i) - x_i),$$

is non-positive because c is a stable fixed point (i.e., $(x_i - c)(s(x_i) - x_i) \leq 0$ for each $x_i \in I$).

The second term,

$$V_{c,2}(\boldsymbol{x}) = -\frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x_i - x_j)(s(x_i) - s(x_j)),$$

is also non-positive because s is non-decreasing and $a_{ij} \geq 0$. Therefore, $\dot{V}_c(\boldsymbol{x}) \leq 0$ in a neighborhood of $\boldsymbol{x}^*$, proving local stability by Lyapunov's theorem 2.4.34.

(ii) If, in addition, c is an isolated stable fixed point, then for $\boldsymbol{x}$ near $c\mathbf{1}$ but not equal to it, at least one term in the sums for $\dot{V}_c(\boldsymbol{x})$ will be strictly negative. Specifically, if $\boldsymbol{x} \neq c\mathbf{1}$, there must exist at least one $x_i \neq c$. Then $(x_i - c)(s(x_i) - x_i) < 0$ for such an $x_i \in I$ because c is an isolated stable fixed point. Thus $\dot{V}_c(\boldsymbol{x}) < 0$ for $\boldsymbol{x} \neq c\mathbf{1}$ in a neighborhood of $c\mathbf{1}$. By LaSalle's Invariance Principle, trajectories must converge to the largest invariant set where $\dot{V}_c(\boldsymbol{x}) = 0$, which is just $\{c\mathbf{1}\}$. This proves local asymptotic stability. The converse follows from the same argument as in the necessity proof for (i).

(iii) This follows directly from the arguments above. If c is an unstable fixed point ($c \in \mathrm{Fix}^\circ(s)$), then by Definition 3.2.3, there exists x in any neighborhood of c such that $(x - c)(s(x) - x) > 0$. This implies that for points $\boldsymbol{x}$ sufficiently close to $c\mathbf{1}$, we can find initial conditions such that $\dot{V}_c(\boldsymbol{x}) > 0$. Thus, trajectories will move away from $\boldsymbol{x}^*$, rendering it unstable. This is consistent with the conditions for instability in one-dimensional dynamics (Theorem 3.5.6-(iii)). □

Theorem 4.3.1 provides a strong and final answer to the question of *local* stability of FSE. It confirms that stability is never lost when moving from the 1D manifold to the full N-dimensional space.

However, local stability is only part of the story. In systems with multiple equilibria, a question remains: what is the *region of convergence*, or *basin of attraction*, for a given stable equilibrium? Answering this requires us to move beyond an only local analysis and characterize the basins of attraction on a larger scale.

4.3.2 Characterization of Basin of Attraction for any FSE

While the previous theorem characterizes the behavior of trajectories near a synchronized equilibrium, we now seek to understand what is happening in the invariant hypercubes defined in Section 3.6 of the previous chapter. Specifically, for a given FSE, what is the set of initial conditions that are guaranteed to converge to it? The following proposition extends the one-dimensional basin of attraction analysis from Chapter 3 (Proposition 3.5.9) to the full N-dimensional system. It reveals that the unstable fixed points of s act as boundaries, defining large hyper-rectangular regions of convergence.

Proposition 4.3.2 (Basin of Attraction of FSE)**.** *Let c^* be a fixed point of s, defining the FSE $\boldsymbol{x}^* = c^*\mathbf{1}$. Let its basin of attraction be $\mathcal{B}(\boldsymbol{x}^*) := \{\boldsymbol{x}(0) \in \mathcal{X} \mid \lim_{t\to\infty} \boldsymbol{x}(t) = \boldsymbol{x}^*\}$. Define the nearest non-stable fixed points bounding c^* as:*

$$\underline{c} := \max(\{c \in \mathrm{Fix}^\circ(s) \mid c < c^*\} \cup \{-1\}) \quad \textit{and} \quad \overline{c} := \min(\{c \in \mathrm{Fix}^\circ(s) \mid c > c^*\} \cup \{1\}).$$

Then the following equivalences hold:

(i) The hypercube below $c^*\mathbf{1}$, $\mathcal{H}_{\mathrm{B}} := [\underline{c}, c^*]^N \setminus \{\underline{c}\mathbf{1}\}$ is a subset of $\mathcal{B}(\boldsymbol{x}^*)$ if and only if $s(x) > x$, for all $x \in (\underline{c}, c^*)$, i.e., $\underline{c} \in \mathrm{Fix}^\circ_{\mathrm{R}}(s) \cup \{-1\}$.

(ii) The region $[c^*, \overline{c}]^N \setminus \{\overline{c}\mathbf{1}\}$ is a subset of $\mathcal{B}(\boldsymbol{x}^*)$ if and only if $s(x) < x$, for all $x \in (c^*, \overline{c})$, i.e., $\overline{c} \in \mathrm{Fix}^\circ_{\mathrm{L}}(s) \cup \{1\}$.

(iii) The region $[\underline{c}, \overline{c}]^N \setminus \{\underline{c}\mathbf{1}, \overline{c}\mathbf{1}\} \subseteq \mathcal{B}(\boldsymbol{x}^*)$ if and only if $\boldsymbol{x}^*$ is locally asymptotically stable. i.e., $c^* \in \mathrm{Fix}^\bullet(s)$.

Those results extend to intervals over $\mathcal{S}$ of fixed point $I \subseteq \mathrm{Fix}(x)$.

Proof. Note that $\underline{c}\mathbf{1}$ and $\overline{c}\mathbf{1}$ are also equilibria of (COSA). Thus, they cannot be in the attraction basin of $\boldsymbol{x}^*$. i.e., $\underline{c}\mathbf{1} \notin \mathcal{B}(x^*)$ and $\overline{c}\mathbf{1} \notin \mathcal{B}(x^*)$.

(i) ($\Rightarrow$) Let us restrict to the FS case since $\mathcal{S}$ is forward invariant. Over, $\mathcal{S}$ we have a 1-dimensional dynamics: $\dot{x} = s(x) - x := f(x)$ defined over $[-1, 1]$.

Since $[\underline{c}, c^*]^N \setminus \{\underline{c}\mathbf{1}\} \subseteq \mathcal{B}(x^*)$, one has $[\underline{c}, c^*]^N \cap \mathcal{S} \setminus \{\underline{c}\mathbf{1}\} \subseteq \mathcal{B}(x^*)$, then for all $x(0) \in (\underline{c}, c^*]$, we must have $\lim_{t\to\infty} x(t) = x^*$. Moreover, since for all $x \in (\underline{c}, c^*)$, we have $x < c^*$, the function $t \mapsto x(t)$ must be increasing. Therefore, $\dot{x}(t) > 0$ for all $t \geq 0$. Then,

$$\dot{x} > 0 \Leftrightarrow s(x) - x > 0 \Leftrightarrow s(x) > x.$$

This yields $s(x) > x$ for all $x \in (\underline{c}, c^*)$.

($\Leftarrow$) Let us consider the function $V(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^\top \boldsymbol{D}(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0$. The time derivative of V along the trajectories of the dynamics is given by

$$\begin{aligned}\dot{V}(\boldsymbol{x}) &= (\boldsymbol{x} - \boldsymbol{x}^*)^\top \boldsymbol{D}\dot{\boldsymbol{x}} = (\boldsymbol{x} - \boldsymbol{x}^*)^\top \boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - (\boldsymbol{x} - \boldsymbol{x}^*)^\top \boldsymbol{D}\boldsymbol{x} \\ &= (\boldsymbol{x} - c^*\mathbf{1})^\top \boldsymbol{D}(\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}) - (\boldsymbol{x} - c^*\mathbf{1})^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}),\end{aligned}$$

using the definition of the Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$. Since $\mathbf{1}^\top$ is a left eigenvector of $\boldsymbol{L}$ associated with eigenvalue 0, this rewrites in summation form:

$$\dot{V}(\boldsymbol{x}) = \sum_{i=1}^{N} d_i(x_i - c^*)(s(x_i) - x_i) - \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x_i - x_j)(s(x_i) - s(x_j)).$$

To use LaSalle's Invariance Principle, we need to show that

$$E = \{\boldsymbol{x} \in \mathcal{X} \mid \dot{V}(\boldsymbol{x}) = 0\} \cap ([\underline{c}, c^*]^N \setminus \{\underline{c}\mathbf{1}\}) = \{\boldsymbol{x}^*\}.$$

This is true since $\boldsymbol{L}\mathbf{1} = 0$ and c^* is the only fixed point of s in $(\underline{c}, c^*]$ (since $s(x) > x$, for all $x \in (\underline{c}, c^*)$ by definition). Additionally, over the hypercube $[\underline{c}, c^*)^N \setminus \{\underline{c}\mathbf{1}\}$, the first term of $\dot{V}(\boldsymbol{x})$ is negative since there exists at least one $i \in \mathcal{V}$ such that $x_i \neq c^*_{\mathrm{L}}$ and then verify $(x_i - c^*)(s(x_i) - x_i) < 0$ by the assumption $s(x_i) > x_i$ and $\sum_{i,j=1}^{N} a_{ij}(x_i - x_j)(s(x_i) - s(x_j)) \geq 0$ by non-decreasing-ness of s. Therefore, V is a Lyapunov function for the dynamics (COSA) over $[\underline{c}, c^*]^N \setminus \{\underline{c}\mathbf{1}\}$.

Let $n_0 \in \mathbb{N}$ such that $[\underline{c} + 1/n_0, c^*] \subset (\underline{c}, c^*]$. For any $i \in \mathcal{V}$, let us note $H_n^i = [\underline{c}, c^*]^{i-1} \times [\underline{c} + 1/n, c^*] \times [\underline{c}, c^*]^{N-i}$ for $n \geq n_0$, this hypercube is a compact forward invariant for the dynamics (COSA) from Proposition 3.6.3-(ii). Additionally, one has $\dot{V}(\boldsymbol{x}) < 0$ for any $\boldsymbol{x} \in H_n^i \setminus \{c^*\mathbf{1}\}$. Thus, by LaSalle's Invariance Principle, one has $H_n^i \subseteq \mathcal{B}(\boldsymbol{x}^*)$. Then, since this is true for all $i \in \mathcal{V}$ and $n \geq n_0$, we have that every $H_n^i \subseteq \mathcal{B}(x^*)$ and thus

$$\bigcup_{n \geq n_0} \bigcup_{i \in \mathcal{V}} H_n^i \subseteq \mathcal{B}(x^*) \Leftrightarrow [\underline{c}, x^*]^N \setminus \{c_{\mathrm{L}^*}\mathbf{1}\} \subseteq \mathcal{B}(x^*).$$

(ii) The proof follows by symmetry from the previous.

(iii) This is a corollary of (i) and (ii).

The extension to intervals over $\mathcal{S}$ uses the same argument except that in that case $\{\boldsymbol{x} \in \mathcal{X} \mid \dot{V}(\boldsymbol{x}) = 0\} = I^N \cap \mathcal{S}$ and is not reduced to a singleton. The results follow from LaSalle's Invariance Principle. □

The preceding analysis demonstrates that the basins of attraction for synchronized equilibria are delimited by the unstable fixed points of s. The following proposition strengthens this by establishing an explicit rate of convergence under a stricter, Lipschitz-like condition on the signal function.

Proposition 4.3.3 (Exponential Convergence Rate to an FSE). *Let c^* be an asymptotically stable fixed point of s and let $\mathcal{H} = [a, b]^N$ be an invariant hypercube within its basin of attraction, such that $c^* \in (a, b)$. Suppose there exists a constant $\alpha \in [0, 1)$ such that the secant slope relative to the equilibrium c^* is uniformly bounded for all $x \in [a, b]$:*

$$0 \leq \frac{s(x) - s(c^*)}{x - c^*} \leq \alpha \quad \text{for all } x \in [a, b] \setminus \{c^*\}. \tag{4.6}$$

Then, any trajectory starting in $\mathcal{H}$ converges to the FSE $\boldsymbol{x}^* = c^*\mathbf{1}$ exponentially. Specifically, the Lyapunov function $V_{c^*}(\boldsymbol{x})$ defined in (4.5) decays according to the bound:

$$V_{c^*}(\boldsymbol{x}(t)) \leq V_{c^*}(\boldsymbol{x}(0))e^{-2(1-\alpha)t}.$$

Proof. We consider the Lyapunov function $V_{c^*}(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - c^*\mathbf{1})^\top \boldsymbol{D}(\boldsymbol{x} - c^*\mathbf{1})$. Its time derivative, derived in the proof of Theorem 4.3.1, is

$$\dot{V}_{c^*}(\boldsymbol{x}) = \underbrace{\sum_{i=1}^{N} d_i(x_i - c^*)(s(x_i) - x_i)}_{V_{c^*,1}(\boldsymbol{x})} \underbrace{- \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x_i - x_j)(s(x_i) - s(x_j))}_{V_{c^*,2}(\boldsymbol{x})}.$$

By the invariance of $\mathcal{H}$, any trajectory starting in it remains within it. For the first term, we rewrite $s(x_i) - x_i = (s(x_i) - s(c^*)) - (x_i - c^*)$. The secant condition (4.6) directly implies $s(x_i) - s(c^*) = \gamma_i(x_i - c^*)$ for some $\gamma_i \in [0, \alpha]$. Thus, $s(x_i) - x_i = (\gamma_i - 1)(x_i - c^*)$, and the first term is bounded as:

$$V_{c^*,1}(\boldsymbol{x}) = \sum_{i=1}^{N} d_i(x_i - c^*)(\gamma_i - 1)(x_i - c^*) = \sum_{i=1}^{N} d_i(\gamma_i - 1)(x_i - c^*)^2 \leq (\alpha - 1)\sum_{i=1}^{N} d_i(x_i - c^*)^2.$$

For the second term $V_{c^*,2}(\boldsymbol{x})$, condition (4.6) implies that s is non-decreasing on $[a, b]$, so $(x_i - x_j)(s(x_i) - s(x_j)) \geq 0$. Since $a_{ij} \geq 0$, it follows that $V_{c^*,2}(\boldsymbol{x}) \geq 0$.

Combining these bounds yields the differential inequality:

$$\dot{V}_{c^*}(\boldsymbol{x}) = V_{c^*,1}(\boldsymbol{x}) - V_{c^*,2}(\boldsymbol{x}) \leq (\alpha - 1)\sum_{i=1}^{N} d_i(x_i - c^*)^2 = 2(\alpha - 1)V_{c^*}(\boldsymbol{x}).$$

Since $\alpha < 1$, the constant $k = 2(1 - \alpha)$ is positive. By the Comparison Lemma, this implies the exponential decay $V_{c^*}(\boldsymbol{x}(t)) \leq V_{c^*}(\boldsymbol{x}(0))e^{-kt}$, proving the proposition. □

While Proposition 4.3.3 guarantees exponential convergence to a specific FSE, a stronger condition on the signal function ensures that the entire synchronization manifold is exponentially attractive. This result is important for understanding global synchronization, as it guarantees that disagreement among agents is suppressed at an exponential rate, irrespective of the evolution of the weighted average opinion.

Proposition 4.3.4 (Exponential Attraction of the Synchronization Manifold)**.** *Let $\mathcal{H} = [a, b]^N$ be an invariant hypercube contained in the basin of attraction of the synchronized manifold $\mathcal{S}$. Suppose there exist constants β and α such that for all distinct $x, y \in [a, b]$, the secant slope is bounded as:*

$$0 < \beta \leq \frac{s(x) - s(y)}{x - y} \leq \alpha < 1. \tag{4.7}$$

Then, the fully synchronized manifold $\mathcal{S}$ is exponentially attractive for all trajectories within $\mathcal{H}$. Specifically, the disagreement measure $V(\boldsymbol{x})$ from (4.4) decays according to the bound:

$$V(\boldsymbol{x}(t)) \leq V(\boldsymbol{x}(0))e^{-k_\perp t},$$

where the rate constant $k_\perp = 2(1 - \alpha) + 2\beta\lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L})$ depends on both the signal function and the graph's algebraic connectivity (Definition 2.2.18).

Proof. We analyze the time derivative of the disagreement measure $V(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{e}$, where $\boldsymbol{e} = \boldsymbol{x} - \bar{x}\mathbf{1}$ is the synchronization error vector from Definition 4.2.3.

The derivative is $\dot{V}(\boldsymbol{x}) = \boldsymbol{e}^\top \boldsymbol{D}\dot{\boldsymbol{e}}$. Since $\dot{\boldsymbol{e}} = \dot{\boldsymbol{x}} - \dot{\bar{x}}\mathbf{1}$ and $\boldsymbol{e}$ is $\boldsymbol{D}$-orthogonal to $\mathbf{1}$ by Lemma 4.2.4, we have $\dot{V}(\boldsymbol{x}) = \boldsymbol{e}^\top \boldsymbol{D}\dot{\boldsymbol{x}}$. Substituting the system dynamics (COSA) and using $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ yields:

$$\dot{V}(\boldsymbol{x}) = \boldsymbol{e}^\top \boldsymbol{D}(\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}) = \boldsymbol{e}^\top \boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{x}.$$

Since $\boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{x} = \boldsymbol{e}^\top \boldsymbol{D}(\boldsymbol{e} + \bar{x}\mathbf{1}) = \boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{e} = 2V(\boldsymbol{x})$, this becomes:

$$\dot{V}(\boldsymbol{x}) = \boldsymbol{e}^\top(\boldsymbol{D} - \boldsymbol{L})\boldsymbol{s}(\boldsymbol{x}) - 2V(\boldsymbol{x}) = \underbrace{\boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{s}(\boldsymbol{x})}_{V_1(\boldsymbol{x})} - \underbrace{\boldsymbol{e}^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x})}_{V_2(\boldsymbol{x})} - 2V(\boldsymbol{x}).$$

Let us bound the two terms separately. For $V_1(\boldsymbol{x})$, using the orthogonality property $\sum d_i e_i = 0$, we can center the signal function at the weighted average $\bar{x} \in [a, b]$:

$$V_1(\boldsymbol{x}) = \sum_{i=1}^{N} d_i e_i s(x_i) = \sum_{i=1}^{N} d_i e_i (s(x_i) - s(\bar{x})).$$

The upper bound in (4.7) implies $s(x_i) - s(\bar{x}) = \gamma_i(x_i - \bar{x}) = \gamma_i e_i$ for some $\gamma_i \in [\beta, \alpha]$. Thus, $V_1(\boldsymbol{x})$ can be bounded as:

$$V_1(\boldsymbol{x}) = \sum_{i=1}^{N} d_i \gamma_i e_i^2 \leq \alpha \sum_{i=1}^{N} d_i e_i^2 = \alpha(\boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{e}) = 2\alpha V(\boldsymbol{x}).$$

For $V_2(\boldsymbol{x})$, this term can be expressed as the Laplacian quadratic form as in the proof of Theorem 4.3.1

$$V_2(\boldsymbol{x}) = \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x_i - x_j)(s(x_i) - s(x_j)).$$

Using the lower bound in (4.7), we have:

$$V_2(\boldsymbol{x}) \geq \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x_i - x_j)(\beta(x_i - x_j)) = \beta(\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}).$$

Since $\boldsymbol{L}\mathbf{1} = \mathbf{0}$, we have $\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \boldsymbol{e}^\top \boldsymbol{L}\boldsymbol{e}$. Let $\nu_2 = \lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L})$ be the second smallest eigenvalue of the normalized Laplacian. From the Courant-Fischer theorem [115, Theorem 4.2.6], we know that for any vector $\boldsymbol{z}$ that is $\boldsymbol{D}$-orthogonal to $\mathbf{1}$, $\boldsymbol{z}^\top \boldsymbol{L}\boldsymbol{z} \geq \nu_2 \boldsymbol{z}^\top \boldsymbol{D}\boldsymbol{z}$. Since $\boldsymbol{e}$ satisfies this condition, we have:

$$\boldsymbol{e}^\top \boldsymbol{L}\boldsymbol{e} \geq \nu_2 \boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{e} = 2\nu_2 V(\boldsymbol{x}).$$

Combining these results gives the final bound for $V_2(\boldsymbol{x})$:

$$V_2(\boldsymbol{x}) \geq 2\beta\lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L})V(\boldsymbol{x}).$$

Substituting the bounds for $V_1(\boldsymbol{x})$ and $V_2(\boldsymbol{x})$ into the expression for $\dot{V}(\boldsymbol{x})$:

$$\dot{V}(\boldsymbol{x}) \leq 2\alpha V(\boldsymbol{x}) - 2\beta\lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L})V(\boldsymbol{x}) - 2V(\boldsymbol{x}) = -\big(2(1-\alpha) + 2\beta\lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L})\big)V(\boldsymbol{x}).$$

With $k_\perp = 2(1-\alpha) + 2\beta\lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L})$, we have the differential inequality $\dot{V}(\boldsymbol{x}) \leq -k_\perp V(\boldsymbol{x})$. Since $\alpha < 1$, $\beta > 0$, and $\lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L}) > 0$ for a connected graph, $k_\perp > 0$. The Gronwall's Inequality implies that $V(\boldsymbol{x}(t)) \leq V(\boldsymbol{x}(0))e^{-k_\perp t}$, which proves the exponential attraction of the manifold $\mathcal{S}$ with rate $k_\perp$. □

[115] *Matrix Analysis*. Roger A. Horn et al. (2017).

Remark 4.3.1 (On the Domain of Validity for Exponential Rates)**.** *The secant conditions in the preceding propositions are key to establishing an exponential rate of convergence. It is needed to understand their domain of validity.*

For Proposition 4.3.3, the condition relies on the secant slope relative to the stable fixed point c^*. A direct consequence of the stability of c^* is that the inequality $\frac{s(x)-s(c^*)}{x-c^*} < 1$ holds for every point x within the open basin of attraction $(\underline{c}, \overline{c})$. While this guarantees convergence, it is not sufficient for a uniform exponential rate *over the whole basin. The ratio, while always less than 1, approaches 1 as x nears the unstable boundaries $\underline{c}$ or $\overline{c}$.*

Therefore, for both propositions, a uniform bound $\alpha < 1$ can only be found on a compact sub-hypercube $\mathcal{H} = [a, b]^N$ strictly inside the basin. These results correctly establish that for any such region, there is a guaranteed minimum exponential rate of convergence. This rate naturally approaches zero as the chosen region expands to fill the entire basin, illustrating the fact that trajectories starting near the unstable boundaries converge arbitrarily slowly.

Remark 4.3.2. *Proposition 4.3.4 provides an intuitive decomposition of the factors that govern the rate of synchronization within the sub-hypercube $\mathcal{H} = [a, b]^N$ strictly contained in the basin of attraction. The rate constant, $k_\perp = 2(1-\alpha) + 2\beta\lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L})$, is the sum of two distinct contributions representing the interplay between local agent dynamics and the network's topological structure.*

The first term, $2(1-\alpha)$, can be interpreted as a ***local contraction rate****. It arises from the fact that the signal function $s(x)$ is less steep than the identity function (since $\alpha < 1$). This property ensures that the local dynamics, $\dot{x}_i = s(x_i) - x_i$, are inherently stable around the equilibrium line $y = x$, pulling each agent's state towards it independently of network interactions. A smaller α implies a stronger local pull towards equilibrium, hence a faster rate contribution.*

The second term, $2\beta\lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L})$, represents the ***network synchronization rate****. This term establishes a direct connection to the classical linear consensus literature. The algebraic connectivity of the normalized Laplacian, $\lambda_2(\boldsymbol{D}^{-1}\boldsymbol{L})$, is precisely the exponential convergence rate for the disagreement vector in the standard linear consensus dynamics $\dot{\boldsymbol{x}} = -\boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{x}$. In our nonlinear model, the constant β acts as a lower bound on the effective "gain" of the network interactions. It quantifies the minimum extent to which the signal function amplifies differences between connected agents, thereby driving them together.*

Thus, the overall rate $k_\perp$ combines the stabilizing influence of the local dynamics with the cohesive strength of the network topology, with the latter's effect being scaled by the sensitivity of the signal function. Synchronization is guaranteed to be fast if either the local dynamics are strongly contractive (small α) or the network is well-connected (large λ_2) and the signal function is sufficiently responsive (large β).

The preceding propositions establish that when the signal function is sufficiently contractive within a given basin of attraction, convergence is not only guaranteed but also exponentially fast. These results provide a quantitative understanding of the system's behavior within these well-defined regions.

This naturally leads to the question of global behavior: under what conditions does the basin of attraction for the synchronized manifold contains the entire state space? According to Proposition 4.3.2, the boundaries of these basins are determined by unstable fixed points. Therefore, global convergence to the set of FSEs can be expected for classes of signal functions that structurally preclude the existence of such unstable points. In this scenario, the boundaries $\underline{c}$ and $\overline{c}$ effectively become -1 and 1, respectively, making the entire state space $\mathcal{X}$ a single region of convergence.

We now turn our analysis to a prominent class of such functions: underestimating signal

functions. As we will demonstrate, the simple condition $s(x) \leq x$ is sufficient to eliminate any repelling fixed points and guarantee global convergence to the synchronized manifold.

4.3.3 The case of underestimation and single crossing function

The following proposition analyzes the case where the function s is an underestimation.

Proposition 4.3.5 (Underestimation implies convergence to FSM)**.** *Let the signal function s be a underestimation. Then, every solution of* (COSA) *starting in $\mathcal{X}$ approaches a FSE at a fixed point of s. i.e., $\lim_{t\to\infty} \boldsymbol{x}(t) = \boldsymbol{x}^* \in \mathcal{S}[\mathrm{Fix}(s)]$ for all $\boldsymbol{x}(0) \in \mathcal{X}$.*

Proof. This proof is similar to the proof of Theorem 4.3.1-(i), considering the natural fixed point $c = 0$ of s but conclude using LaSalle's Invariance Principle. Consider the Lyapunov candidate function $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ defined by

$$V(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{x} = V_0(\boldsymbol{x}).$$

Calculating the time derivative of V along the trajectories of system (COSA), yields

$$\begin{aligned}\dot{V}(\boldsymbol{x}) &= \boldsymbol{x}^\top \boldsymbol{D}\dot{\boldsymbol{x}} = \boldsymbol{x}^\top \boldsymbol{D}(\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{s}(\boldsymbol{x}) + \boldsymbol{x}^\top \boldsymbol{D}(\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}) \\ &= -\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}) + \boldsymbol{x}^\top \boldsymbol{D}(\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}),\end{aligned}$$

where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ is the Laplacian matrix. As in proof of Theorem 4.3.1-(i), both terms are non-positive.

Let us now characterize the largest invariant set $E = \{\boldsymbol{x} \in \mathcal{X} \mid \dot{V}(\boldsymbol{x}) = 0\}$. The condition $\dot{V}(\boldsymbol{x}) = 0$ is satisfied if and only if both terms in the derivative vanish simultaneously. The first term $-\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{x}) = 0$ implies that $x_i = x_j$ for all $i, j \in \mathcal{V}$, while the second term $\boldsymbol{x}^\top \boldsymbol{D}\left(\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}\right) = 0$ implies that $x_i = s(x_i)$ for all $i \in \mathcal{V}$. The first condition establishes that $\boldsymbol{x} \in \mathcal{S}$, which by Proposition 3.5.2 is forward invariant. Therefore, the largest invariant set satisfying $\dot{V}(\boldsymbol{x}) = 0$ can be characterized as $\mathcal{S} \cap \mathrm{Fix}(s)^N$ which is precisely the set $\mathcal{S}[\mathrm{Fix}(s)]$.

The convergence of solutions follows from LaSalle's Invariance Principle. Given that $\mathcal{X}$ is compact and forward invariant, $V(\boldsymbol{x})$ is bounded from below on $\mathcal{X}$. Moreover, since $\dot{V}(\boldsymbol{x}) \leq 0$, and $\mathcal{S}[\mathrm{Fix}(s)]$ is the largest invariant set where $\dot{V}(\boldsymbol{x}) = 0$, we can conclude that every solution with initial condition in $\mathcal{X}$ converges to $\mathcal{S}[\mathrm{Fix}(s)]$ as $t \to \infty$. □

This proposition establishes a sufficient condition for global synchronization. It formalizes the intuition that if the communication consistently attenuating the expression of states relative to the real ones (i.e., underestimation), disagreement is not viable outcome. The cooperative network interactions then ensure that the system inevitably settles into a state of complete consensus, regardless of the communication topology. This behavior is illustrated in Figure 4.3 for a instrumental signal function.

The principle underlying this global convergence result extends beyond strictly underestimating functions. The central feature is the absence of any "repelling" fixed points that could sustain disagreement. This leads to a natural generalization for any signal function that intersects the identity line precisely once, a property we term *single-crossing* (term used in economics literature).

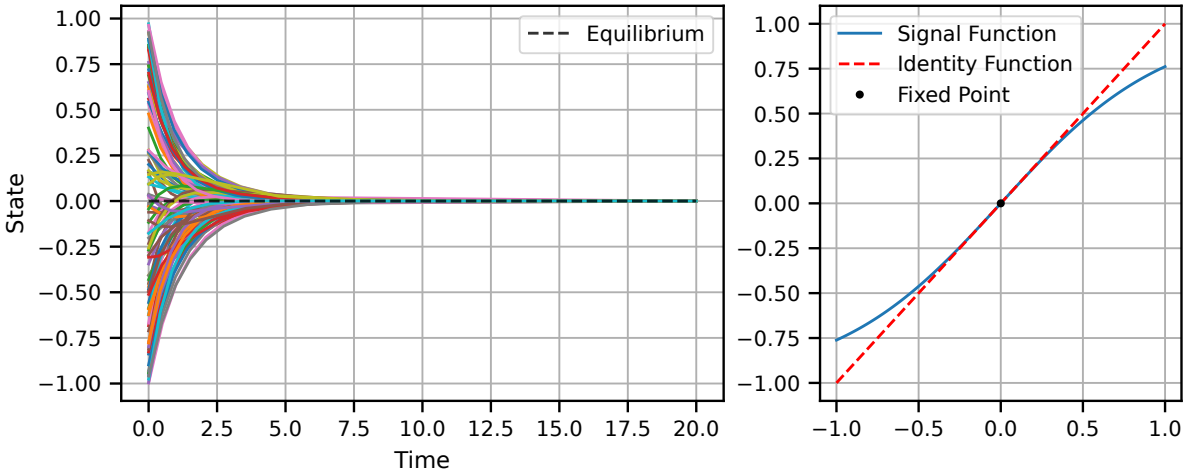


Figure 4.3: State evolution for agents with global underestimation signal over a connected graph (Erdős-Rényi with $N = 100$ and $p = 0.1$) with random uniform initial conditions and $s(x) = \tanh(x)$.

Corollary 4.3.6 (Single-crossing implies convergence to crossing point)**.** *Let the signal function s be a single-crossing function with crossing point c. Then, every solution of* (COSA) *starting in $\mathcal{X}$ approaches the FSE at c. i.e., $\lim_{t\to\infty} \boldsymbol{x}(t) = c\mathbf{1}$ for all $\boldsymbol{x}(0) \in \mathcal{X}$.*

Proof. The proof is a direct consequence of Proposition 4.3.5. In this case, the set $\mathrm{Fix}^{\bullet}(s) = \{c\}$ is a singleton. Thus, the basin of attraction is the entire state space $\mathcal{X}$. □

Together, Proposition 4.3.5 and its Corollary 4.3.6 identify a class of systems where global synchronization is guaranteed based only on the properties of the agent's signal function. When $s(x)$ is either an underestimation or single-crossing, the dynamics lack the repelling equilibria necessary to build disagreement. Consequently, the cooperative nature of the interactions drives the entire network to a consensus state, irrespective of the graph's specific structure. This hints that the shape of the signal function plays an important roles in the behavior of the system. We will explicitly state this role in the next section.

4.4 The Interplay of Signal Function and Network Topology

The preceding analysis has established a broad class of systems for which global synchronization is guaranteed, based solely on properties of the signal function s having non unstable fixed point (i.e., $\mathrm{Fix}^{\circ}(s) = \emptyset$). This naturally leads to plenty of natural questions:

- What happens when these straightforward conditions on s are not met?
- If a NFSE exists, what are the core properties s must possess, and where in the state space can it reside?
- Can the network topology overcome this divisive potential of s?

This section analyzes the interplay between these two forces in the aim of answering those questions. We begin by establishing necessary conditions that any NFSE must satisfy, revealing that disagreement is a highly structured phenomenon build upon the unstable fixed points of

s. We then explore two extremes of this interplay: first, topologies so cohesive that they enforce consensus for *any* admissible signal function, and second, signal functions so nonlinear that they can induce disagreement on *any* non-trivial network. This investigation extends deeper into the next section, where we will derive the sharp threshold that precisely quantifies this balance.

4.4.1 Necessary Conditions for the Emergence of Disagreement

Theorem 4.4.1 (Necessary Conditions for an NFSE). *Let $\boldsymbol{x}^*$ be an NFSE of the dynamics (COSA). i.e., $\boldsymbol{x}^*$ is an equilibrium that lies outside the fully synchronized manifold $\mathcal{S}$. Then:*

(i) The set of non-stable fixed points must be non-empty, i.e., $\mathrm{Fix}^\circ(s) \neq \emptyset$.

(ii) There must exist a non-stable fixed point $c^\circ \in \mathrm{Fix}^\circ(s)$ that splits the agents into two non-empty sets, $\mathcal{I} = \{i \mid x_i^* < c^\circ\}$ and $\mathcal{J} = \{j \mid x_j^* > c^\circ\}$. Furthermore, both of these sets must contain at least two agents: $|\mathcal{I}| \geq 2$ and $|\mathcal{J}| \geq 2$.

(iii) Let $x_m^* = \min_{i\in\mathcal{V}} x_i^*$ and $x_M^* = \max_{i\in\mathcal{V}} x_i^*$. The signal function s must possess both a left-unstable fixed point, c_{L}°, and a right-unstable fixed point, c_{R}°, that lie strictly between the extremal states of the equilibrium: $x_m^* < c_{\mathrm{L}}^\circ \leq c_{\mathrm{R}}^\circ < x_M^*$.

Proof. (i) We will do a proof by contradiction. Assume there exists a NFSE $\boldsymbol{x}^* \notin \mathcal{S}$ while $\mathrm{Fix}^\circ(s)$ is empty. If $\mathrm{Fix}^\circ(s)$ is empty, then s does not reach the identity line from below. Due to the continuity of s, this implies there is either a unique fixed point $c^\bullet$ of s or a continuum of them (i.e., there exists an interval $I \subset [-1, 1]$ over which s is an exact estimation). The same reasoning applies to both situations by taking $c^\bullet$ as the minimal fixed point in the case of a continuum. Then, $x \mapsto s(x - c^\bullet) - c^\bullet$ is an underestimation. By Proposition 4.3.5, if s is an underestimation over $[-1, 1]$, it implies that $\mathcal{S}$ is globally asymptotically stable, thus there is no non-synchronized equilibrium contradicting the premise.

(ii) Let $\boldsymbol{x}^* \notin \mathcal{S}$ be given. Let $x_m^* = \min_{i\in\mathcal{V}} x_i^*$ and $x_M^* = \max_{i\in\mathcal{V}} x_i^*$. We first show by contradiction that there must exist some $c^\circ \in \mathrm{Fix}^\circ(s)$ such that $x_m^* < c^\circ < x_M^*$.

Suppose that for all $c^\circ \in \mathrm{Fix}^\circ(s)$, $c^\circ \leq x_i^* \leq 1$ for all $i \in \mathcal{V}$. This must be true for their maximum $c_{\max}^\circ = \max(\mathrm{Fix}^\circ(s))$, then $c_{\max}^\circ \leq x_i \leq 1$. By assumption, $\boldsymbol{x}^* \notin \mathcal{S}$, thus we must have $x_M^* > c_{\max}^\circ$ and $x_m < 1$. i.e., $\boldsymbol{x}^* \in [c_{\max}^\circ, 1]^N \setminus \{c_{\max}^\circ \mathbf{1}, \mathbf{1}\}$.

Now there are two possible situations:

- $\mathrm{Fix}^\bullet(s) \cap (c_{\max}^\circ, 1] = \emptyset$. Then, by continuity of s we must have $s(x) \leq x$ for all $x > c_{\max}^\circ$. Since $c_{\max}^\circ$ is the maximum of $\mathrm{Fix}^\circ(s)$, the inequality is strict, otherwise it would exist another unstable fixed point of s greater than $c_{\max}^\circ$. i.e., $s(x) < x$ for all $x > c_{\max}^\circ$in this case, from Proposition 4.3.2-(ii), one has $\boldsymbol{x}^* \in \mathcal{B}(c_{\max}^\circ \mathbf{1})$.
- or, $\mathrm{Fix}^\bullet(s) \cap (c_{\max}^\circ, 1] \neq \emptyset$: If this set is reduced to a singleton $c^\bullet$ in this case by Proposition 4.3.2-(iii) one has $[c_{\max}^\circ, 1]^N \setminus \{c_{\max}^\circ \mathbf{1}, \mathbf{1}\} \subseteq \mathcal{B}(c^\bullet \mathbf{1})$ and implies $\boldsymbol{x}^* \in \mathcal{B}(c^\bullet \mathbf{1})$. If this set is an interval I, the reasoning is the same, using the extension of Proposition 4.3.2-(iii) to intervals and $\boldsymbol{x}^* \in \mathcal{B}(I^N \cap \mathcal{S})$. i.e., $\boldsymbol{x}^*$ is in the attraction basin of a continuum of equilibria.

In both cases, $\boldsymbol{x}^*$ is in the attraction basin of FSE equilibria. Moreover, $\boldsymbol{x}^* \in \mathcal{B}(\boldsymbol{x}^*)$ by definition. Since $\boldsymbol{x}^*$ cannot be in two different attraction basin, this contradicts the assumption that for all $c^\circ \in \mathrm{Fix}^\circ(s)$, $x_i^* \geq c^\circ$ for all $i \in \mathcal{V}$. Thus, there exists $c^\circ \in \mathrm{Fix}^\circ(s)$ such that $x_m^* < c^\circ < x_M^*$.

Finally by the equilibrium condition (3.10), with the c° that we have found, at equilibrium we have :

$$x_m^* = \frac{1}{d_m} \sum_{j=1}^{N} a_{mj} s(x_j^*) < c^\circ$$

Since x_m^* is a convex combination of the family $(s(x_i^*))_{i\in\mathcal{N}_m}$, this implies there must exist some $j \in \mathcal{N}_m$ with $s(x_j^*) < c^\circ$. Then by non-decreasing-ness of s on has $s(x_m) \leq s(x_j) < c^\circ$. Define $\mathcal{I} = \{i \in \mathcal{V} \mid s(x_i) < c^\circ\}$, clearly, $j, m \in \mathcal{I}$, moreover, since $j \in \mathcal{N}_m$ and $m \notin \mathcal{N}_m$ (since there is self loop in the graph $\mathcal{G}$), this implies, $j \neq m$ and we conclude that $\mathcal{I}$ has at least two elements. Similarly, since $x_M^* > c^\circ$, the same reasoning apply to show that $s(x_M) > c^\circ$, and that $\mathcal{J} = \{j \in \mathcal{V} \mid s(x_j) > c^\circ\}$ has at least two elements.

(iii) Let $\boldsymbol{x}^*$ be an NFSE, so $x_m^* < x_M^*$. Let $g(x) = s(x) - x$. The equilibrium condition (3.10) give us:

$$x_m^* = \frac{1}{d_m}\sum_{j=1}^{N} a_{mj}s(x_j^*) \geq \frac{1}{d_m}\sum_{j=1}^{N} a_{mj}s(x_m^*) = s(x_m^*).$$

We then have $x_m^* \geq s(x_m^*)$ implying $g(x_m^*) \leq 0$. A symmetric argument yields $g(x_M^*) \geq 0$. If the inequalities are strict, there exist x_m', x_M' with $x_m^* \leq x_m' < x_M' \leq x_M^*$ such that $g(x_m') < 0$ and $g(x_M') > 0$.

Otherwise, if $g(x_m^*) = 0$, since $x_j^* \geq x_m^*$ for all $j \in \mathcal{N}_m$ and s is non-decreasing we must have $s(x_j^*) \geq x_m^*$. Moreover, as x_m^* is a convex combination of $(s(x_j^*))_{j\in\mathcal{N}_m}$, we must have $s(x_j^*) = x_M^*$ for all $j \in \mathcal{N}_m$. Since $\boldsymbol{x}^*$ is an NFSE on a connected graph, there exists $k \in \mathcal{N}_m$ with $x_k^* > x_m^*$ (this statement must be true for at least one $m^* \in \arg\max_{i\in\mathcal{V}} x_i^*$, otherwise it imply $\boldsymbol{x} \in \mathcal{S}$ by propagation of equality to all the graph), implying s is constant on $[x_m^*, x_k^*]$. Thus, for any $x \in (x_m^*, x_k^*]$, $g(x) = x_m^* - x < 0$. Symmetrically, if $g(x_M^*) = 0$, we can find x near and below x_M^* where $g(x) > 0$. Hence, there exist x_m', x_M' with $x_m^* \leq x_m' < x_M' \leq x_M^*$ such that $g(x_m') < 0$ and $g(x_M') > 0$.

Let $c_{\mathrm{L}}^\circ := \inf\{x \in (x_m', x_M') \mid g(x) \geq 0\}$ and $c_{\mathrm{R}}^\circ := \sup\{x \in (x_m', x_M') \mid g(x) \leq 0\}$. By continuity of g, both are fixed points of s. i.e., $g(c_{\mathrm{L}}^\circ) = 0$ and $g(c_{\mathrm{R}}^\circ) = 0$. The construction implies that for some $\delta > 0$, $g(x) < 0$ on $(c_{\mathrm{L}}^\circ - \delta, c_{\mathrm{L}}^\circ)$ and $g(x) > 0$ on $(c_{\mathrm{R}}^\circ, c_{\mathrm{R}}^\circ + \delta)$, so $c_{\mathrm{L}}^\circ \in \mathrm{Fix}_{\mathrm{L}}^\circ(s)$ and $c_{\mathrm{R}}^\circ \in \mathrm{Fix}_{\mathrm{R}}^\circ(s)$. By definition, the strict inequalities $x_m^* < c_{\mathrm{L}}^\circ < x_M^*$ and $x_m^* < c_{\mathrm{R}}^\circ < x_M^*$ hold.

Finally, if $c_{\mathrm{L}}^\circ > c_{\mathrm{R}}^\circ$, then for any $x \in (c_{\mathrm{R}}^\circ, c_{\mathrm{L}}^\circ)$, the definitions would imply both $g(x) > 0$ and $g(x) < 0$, a contradiction. Thus, $c_{\mathrm{L}}^\circ \leq c_{\mathrm{R}}^\circ$, which completes the proof. □

This theorem provides a clear insight: an NFSE can only emerge if the signal function s is sufficiently nonlinear to create "repellers" on the FSM. These unstable fixed points act as pivots, allowing the network forces to balance agents on opposing sides into a state of structured disagreement.

This result immediately yields a powerful and practical corollary for guaranteeing global synchronization based solely on the shape of the signal function.

Corollary 4.4.2 (Sufficient Condition for Global Synchronization via Function Shape)**.** *If the signal function s has no unstable or semi-stable fixed points (i.e., $\mathrm{Fix}^\circ(s) = \emptyset$), then the* (COSA) *dynamics admits no NFSE. Consequently, almost all trajectory of the system converges to the set of synchronized equilibria, $\mathcal{S}[\mathrm{Fix}(s)]$. If additionally s is strictly increasing, the result holds for all trajectories.*

Proof. The proof is a direct application of Theorem 4.4.1-(i) with Proposition 3.3.2 and Proposition 3.3.4. □

The structural requirement from Theorem 4.4.1-(ii), that any disagreement must partition the agents, has a direct consequence on the minimum network size required to support such a state.

Corollary 4.4.3 (Small networks imply FSE)**.** *If the network size $N \leq 3$, then all equilibria of* (COSA) *are FSE.*

Proof. The proof is a direct consequence of Theorem 4.4.1-(ii). □

4.4.2 Synchronization Enforced by Network Topology

Having established with Theorem 4.4.1 the role of the signal function in the emergence of disagreement, we now turn to the role of the network topology. We will consider a class of network for which regardless of the signal function, synchronization is the only asymptotic outcome.

Proposition 4.4.4 (Complete graphs imply FSE)**.** *Let $\mathcal{G}$ be a complete graph. Then, every trajectory of the system* (COSA) *starting from any initial condition $\boldsymbol{x}(0) \in \mathcal{X}$ synchronizes.*

Proof. Consider the variance of the opinions as a Lyapunov function candidate:

$$V(\boldsymbol{x}) = \frac{1}{4}\sum_{i=1}^{N}\sum_{j=1}^{N}(x_i - x_j)^2.$$

This function is non-negative, i.e., $V(\boldsymbol{x}) \geq 0$. Furthermore, $V(\boldsymbol{x}) = 0$ if and only if $x_i = x_j$ for all i, j, which means $V(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x} \in \mathcal{S}$. The time derivative of V along the trajectories of the system:

$$\dot{V}(\boldsymbol{x}) = \frac{1}{2}\sum_{i,j=1}^{N}(x_i - x_j)(\dot{x}_i - \dot{x}_j)$$

Note that $\dot{x}_i - \dot{x}_j$ is given by:

$$\begin{aligned}\dot{x}_i - \dot{x}_j &= \Big[\frac{1}{N-1}\sum_{k=1,k\neq i}^{N} s(x_k) - x_i\Big] - \Big[\frac{1}{N-1}\sum_{k=1,k\neq j}^{N} s(x_k) - x_j\Big] \\ &= \frac{1}{N-1}\Big[\sum_{k=1,k\neq i}^{N} s(x_k) - \sum_{k=1,k\neq j}^{N} s(x_k)\Big] - (x_i - x_j) = \frac{1}{N-1}(s(x_j) - s(x_i)) - (x_i - x_j).\end{aligned}$$

Let us substitute this into the expression for $\dot{V}$:

$$\begin{aligned}\dot{V}(\boldsymbol{x}) &= \frac{1}{2}\sum_{i,j=1}^{N}(x_i - x_j)\left[\frac{1}{N-1}(s(x_j) - s(x_i)) - (x_i - x_j)\right] \\ &= \frac{1}{2(N-1)}\sum_{i,j=1}^{N}(x_i - x_j)(s(x_j) - s(x_i)) - \frac{1}{2}\sum_{i,j=1}^{N}(x_i - x_j)^2 \\ &= -\frac{1}{2(N-1)}\sum_{i,j=1}^{N}(x_i - x_j)(s(x_i) - s(x_j)) - 2V(\boldsymbol{x}).\end{aligned}$$

The function s is non-decreasing, we have $(x_i - x_j)(s(x_i) - s(x_j)) \geq 0$ for all i, j. Consequently, the summation term is non-negative and both terms in the expression for $\dot{V}$ are

non-positive. We can thus conclude that $\dot{V}(\boldsymbol{x}) \leq 0$. This shows that the total disagreement among agents is non-increasing along any trajectory.

By Proposition 3.3.1, the compact set $\mathcal{X}$ is forward invariant. Thus, we can apply LaSalle's Invariance Principle, which implies that every trajectory converges to the largest invariant set $\mathcal{M}$ contained in $E = \{\boldsymbol{x} \in \mathcal{X} \mid \dot{V}(\boldsymbol{x}) = 0\}$.

Let's characterize the set E.

$$\dot{V}(\boldsymbol{x}) = -\frac{1}{2(N-1)} \sum_{i,j=1}^{N} \underbrace{(x_i - x_j)(s(x_i) - s(x_j))}_{\geq 0} - \underbrace{2V(\boldsymbol{x})}_{\geq 0} = 0$$

Since both terms are non-negative, their sum is zero if and only if both terms are individually zero. Since $V(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x} \in \mathcal{S}$, we have, $\{\boldsymbol{x} \in \mathcal{X} \mid \dot{V}(\boldsymbol{x}) = 0\} = \mathcal{S}$. This means the set E is exactly the synchronization manifold $\mathcal{S}$, additionally since $\mathcal{S}$ is itself an invariant manifold (Proposition 3.5.2), it is the largest invariant set within itself. Thus, by LaSalle's Invariance Principle, every trajectory converges to $\mathcal{S}$. □

More generally, connectivity can be quantified by the spectrum of the graph Laplacian as discussed in Chapter Networked Systems Preliminaries in Section 2.2.3 and recalled in Proposition 3.2.1. The following lemma uses spectral graph theory to generalize the complete graph result.

Lemma 4.4.5 (Highly connected graphs imply FSE)**.** *Let $\mathcal{G}$ be such that $\lambda_{N-1} < 0$. Any equilibrium point $\boldsymbol{x}^*$ of* (COSA) *must be a FSE.*

Proof. An equilibrium state $\boldsymbol{x}^*$ must satisfy the equilibria condition (3.9): $\boldsymbol{x}^* = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}^*)$. Let $\boldsymbol{x}^*_\perp$ be the component of $\boldsymbol{x}^*$ in the disagreement subspace $\mathcal{S}^\perp$. On this subspace, the operator $\boldsymbol{D}^{-1}\boldsymbol{A}$ is invertible, as all its eigenvalues $\lambda_1 \leq \cdots \leq \lambda_{N-1}$ are strictly negative and therefore non-zero. Let us note $\boldsymbol{Q} = \boldsymbol{D}^{-1}\boldsymbol{A}$. Applying the equilibrium condition to the disagreement component gives $\boldsymbol{x}^*_\perp = \boldsymbol{Q}\boldsymbol{s}(\boldsymbol{x}^*)_\perp$, which implies $\boldsymbol{s}(\boldsymbol{x}^*)_\perp = \boldsymbol{Q}^{-1}\boldsymbol{x}^*_\perp$.

As in proof of Theorem 4.3.1-(i) since s is non-decreasing we have that for any vector $\boldsymbol{y} \in \mathcal{X}$,

$$\boldsymbol{y}^\top \boldsymbol{L}\boldsymbol{s}(\boldsymbol{y}) = \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(y_i - y_j)(s(y_i) - s(y_j)) \geq 0.$$

Since $\boldsymbol{L} = \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{Q})$, this property can be written in terms of the weighted inner product as $\langle \boldsymbol{y}, (\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{s}(\boldsymbol{y})\rangle_{\boldsymbol{D}} \geq 0$. Applying this inequality at the equilibrium point $\boldsymbol{x}^*$, we have:

$$\langle \boldsymbol{x}^*, (\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{s}(\boldsymbol{x}^*)\rangle_{\boldsymbol{D}} \geq 0.$$

Since $\boldsymbol{I} - \boldsymbol{Q}$ maps the synchronization subspace to zero, only the disagreement components contribute, so:

$$\langle \boldsymbol{x}^*_\perp, (\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{s}(\boldsymbol{x}^*)_\perp\rangle_{\boldsymbol{D}} \geq 0.$$

Substituting $\boldsymbol{s}(\boldsymbol{x}^*)_\perp = \boldsymbol{Q}^{-1}\boldsymbol{x}^*_\perp$:

$$\langle \boldsymbol{x}^*_\perp, (\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{Q}^{-1}\boldsymbol{x}^*_\perp\rangle_{\boldsymbol{D}} = \langle \boldsymbol{x}^*_\perp, (\boldsymbol{Q}^{-1} - \boldsymbol{I})\boldsymbol{x}^*_\perp\rangle_{\boldsymbol{D}} \geq 0.$$

The operator $\boldsymbol{Q}^{-1} - \boldsymbol{I}$ acting on the disagreement subspace $\mathcal{S}^\perp$ has eigenvalues $\{1/\lambda_i - 1\}_{i=1}^{N-1}$. Since by hypothesis $\lambda_i < 0$ for all $i \geq 1$, it follows that $1/\lambda_i < 0$, and thus $1/\lambda_i - 1 < -1$. The operator $\boldsymbol{Q}^{-1} - \boldsymbol{I}$ is therefore strictly negative definite on $\mathcal{S}^\perp$.

The inner product $\langle \boldsymbol{y}, \boldsymbol{M}\boldsymbol{y}\rangle$ where $\boldsymbol{M}$ is a strictly negative definite operator can only be non-negative if $\boldsymbol{y} = \boldsymbol{0}$. Therefore, we must have $\boldsymbol{x}^*_\perp = \boldsymbol{0}$. This means that any equilibrium $\boldsymbol{x}^*$ has no component in the disagreement subspace, and thus must be a synchronized state. □

4.4.3 Disagreement Enabled by Signal Function Nonlinearity

The results in the preceding subsection provide sufficient conditions for synchronization. We now establish a complementary result: a condition under which NFSE are guaranteed to exist for a sufficiently nonlinear signal function. This demonstrates that the interplay between network topology (via λ_{N-1}) and signal nonlinearity (via Lipschitz constant K) is the critical factor.

Proposition 4.4.6 (Existence of NFSE for any Non-trivial Graph)**.** *Let $K > 0$ and $N \geq 4$. For any graph $\mathcal{G}$, such that $\lambda_{N-1} > 0$. The following statements hold:*

(i) If $K\lambda_{N-1} = 1$, the dynamics (COSA) with the specific signal function

$$s(x) = \min(1, \max(-1, Kx))$$

possesses a continuum of NFSE, in $\mathrm{Span}(\boldsymbol{v}_{N-1}) \cap \mathcal{X}$.

(ii) If $K\lambda_{N-1} > 1$, there exists a signal function s verifying Assumption 3.2 (more specifically s is K-Lipschitz) such that dynamics (COSA) possesses NFSE.

Consequently, if $K\lambda_{N-1} \geq 1$, the fully synchronized manifold $\mathcal{S}$ is not globally asymptotically stable for the dynamics (COSA) across all admissible signal functions s.

Proof. (i) By assumption, $\lambda_{N-1} > 0$. Under Assumption 3.1, $\lambda_N = 1$ is simple and we have $\lambda_{N-1} < 1$. Moreover, the eigenvector $\boldsymbol{v}_{N-1}$ cannot be aligned with $\boldsymbol{v}_N = \mathbf{1}$, thus $\boldsymbol{v}_{N-1} \in \mathcal{X} \setminus \mathcal{S}$. Since $K\lambda_{N-1} = 1$ and $\lambda_{N-1} < 1$ one has $K > 1$. Define the signal function $s(x) = \min(1, \max(-1, Kx))$. This function verifies conditions of Assumption 3.2. Consider states of the form $\boldsymbol{x}_\varepsilon = \varepsilon \boldsymbol{v}_{N-1}$ for some scalar $0 < \varepsilon \leq 1/K$. Since $\boldsymbol{x}_\varepsilon \in [-1/K, 1/K]^N$ and the function is linear in this hypercube one has $s(x_i) = Kx_i$ for all $i \in \mathcal{V}$. Therefore, $\boldsymbol{s}(\boldsymbol{x}_\varepsilon) = K\boldsymbol{x}_\varepsilon = K\varepsilon\boldsymbol{v}_{N-1}$. Injecting it into the dynamics yields:

$$\begin{aligned}\dot{\boldsymbol{x}}_\varepsilon &= \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}_\varepsilon) - \boldsymbol{x}_\varepsilon = \varepsilon K \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{v}_{N-1} - \varepsilon\boldsymbol{v}_{N-1} \\ &= \varepsilon(K\lambda_{N-1} - 1)\boldsymbol{v}_{N-1} = \mathbf{0}.\end{aligned}$$

Thus, $\boldsymbol{x}_\varepsilon$ is an NFSE. This being true for all $0 < \varepsilon \leq 1/K$, any $\boldsymbol{x}_\varepsilon$ for such ε is an NFSE. Therefore, there exists a continuum of NFSE in $\mathrm{Span}(\boldsymbol{v}_{N-1}) \cap \mathcal{X}$.

(ii) Assume $K\lambda_{N-1} > 1$. Under Assumption 3.1, one has $\lambda_{N-1} < 1$, which implies $K > 1/\lambda_{N-1} > 1$. We proceed by contradiction. Assume that for *any* signal function s satisfying Assumption 3.2 with Lipschitz constant K, all equilibria of (COSA) are FSE.

We construct a specific signal function s that satisfies these conditions yet leads to a contradiction. Let $s : [-1, 1] \to [-1, 1]$ be a smooth, strictly increasing, and odd function such that $s'(0) = K$. An example of such a function can be constructed based on $s(x) = \tanh(Kx)$ illustrated in Figure 3.3b. Since s is odd and strictly increasing with $s'(0) = K > 1$, its fixed points are $\mathrm{Fix}(s) = \{-c, 0, c\}$ for some $c \in (0, 1]$. By Theorem 4.3.1, the FSE $\boldsymbol{x}^*_{-c} = -c\mathbf{1}$ and $\boldsymbol{x}^*_c = c\mathbf{1}$ are locally asymptotically stable, while the FSE $\mathbf{0}$ is unstable.

Since s is strictly increasing and smooth, by Proposition 3.3.4, every trajectory must converge to one of the three FSE: $-c\mathbf{1}$, $\mathbf{0}$, or $c\mathbf{1}$. This implies a partition of the state space $\mathcal{X}$ into the basins of attraction of these three equilibria:

$$\mathcal{X} = \mathcal{B}(-c\mathbf{1}) \cup \mathcal{B}(\mathbf{0}) \cup \mathcal{B}(c\mathbf{1}).$$

The basins $\mathcal{B}(-c\mathbf{1})$ and $\mathcal{B}(c\mathbf{1})$ are non-empty, disjoint open sets. As the vector field $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}$ is odd due to the oddness of s, the flow is symmetric with respect to the origin,

implying $\mathcal{B}(-c\mathbf{1}) = -\mathcal{B}(c\mathbf{1})$, as observed on the restriction to $\mathcal{S}$ in Figure 3.3b. Consequently, their common boundary, the *separatrix*, is a single set $\partial\mathcal{B}(c\mathbf{1}) = \partial\mathcal{B}(-c\mathbf{1})$.

We now apply [125, Theorem 4.8] to characterize this separatrix. We must verify its assumptions:

(A1) Hyperbolicity: The equilibrium points on the boundary $\partial\mathcal{B}(c\mathbf{1})$ must be hyperbolic.

(A2) Transversality: The stable and unstable manifolds of equilibrium points on the stability boundary satisfy the transversality condition.

(A3) Boundary Convergence: Every trajectory on $\partial\mathcal{B}(c\mathbf{1})$ must approach one of the equilibrium points on the boundary.

Assumption (A2) can be satisfied by a smooth choice of s, as $s(x) = \tanh(Kx)$. To verify (A1) and (A3), we first identify the equilibria on the boundary. Let $\boldsymbol{y} \in \partial\mathcal{B}(c\mathbf{1})$. Since the boundary of an invariant set is itself invariant, the trajectory $\boldsymbol{y}(t)$ remains in $\partial\mathcal{B}(c\mathbf{1})$. As all trajectories converge to an equilibrium, $\boldsymbol{y}(t)$ must converge to one of $\{-c\mathbf{1}, \mathbf{0}, c\mathbf{1}\}$. It cannot converge to $c\mathbf{1}$ (as $\boldsymbol{y}$ would be in the interior of the basin) or $-c\mathbf{1}$ (as the basins are disjoint). Thus, any trajectory on the boundary must converge to $\mathbf{0}$. This implies that $\mathbf{0}$ is the only equilibrium point on the boundary $\partial\mathcal{B}(c\mathbf{1})$, satisfying (A3). The Jacobian at the origin is $\boldsymbol{J}(\mathbf{0}) = s'(0)\boldsymbol{D}^{-1}\boldsymbol{A} - \boldsymbol{I} = K\boldsymbol{D}^{-1}\boldsymbol{A} - \boldsymbol{I}$. Its eigenvalues are $K\lambda_i - 1$, which are non-zero for a generic choice of K, satisfying (A1).

With the assumptions verified, [125, Theorem 4.8 (b)] yields:

$$\partial\mathcal{B}(c\mathbf{1}) = W^s(\mathbf{0}),$$

where $W^s(\mathbf{0})$ is the stable manifold of the origin.

This leads to a topological contradiction based on dimensionality. The set $\partial\mathcal{B}(c\mathbf{1})$ is the boundary separating two open sets in the N-dimensional space $\mathcal{X}$, and thus it must be an $(N-1)$-dimensional manifold. However, the dimension of the stable manifold $W^s(\mathbf{0})$ is equal to the number of eigenvalues of $\boldsymbol{J}(\mathbf{0})$ with negative real parts. The eigenvalues are $\nu_i = K\lambda_i - 1$ for $i = 1, \ldots, N$. By assumption, $K\lambda_{N-1} > 1$. We also know $\lambda_N = 1$, and since $K > 1/\lambda_{N-1} > 1$, we have:

- $\nu_N = K\lambda_N - 1 = K - 1 > 0$.
- $\nu_{N-1} = K\lambda_{N-1} - 1 > 0$.

Thus, at least two eigenvalues of $\boldsymbol{J}(\mathbf{0})$ have positive real parts. The dimension of the unstable manifold, $\dim(W^u(\mathbf{0}))$, is therefore at least 2. Consequently, the dimension of the stable manifold is $\dim(W^s(\mathbf{0})) = N - \dim(W^u(\mathbf{0})) \leq N - 2$.

We then have a contradiction: a manifold of dimension at most $N-2$ cannot separate an N-dimensional space. i.e.,

$$N - 1 = \dim(\partial\mathcal{B}(c\mathbf{1})) = \dim(W^s(\mathbf{0})) \leq N - 2.$$

This contradiction implies that our initial assumption was false. Therefore, for our constructed function s, there must exist at least one equilibrium that is not an FSE. This completes the proof. □

Remark 4.4.1 (Generalization of Sufficiency Condition)**.** *The proof of (ii) of this proposition, establishes that $\lambda_{N-1}K \geq 1$ is a sufficient condition for the existence of a signal function s inducing an NFSE. This result is analogous to the sufficiency result in [126, Theorem 6], which uses bifurcation theory to prove the emergence of mixed-sign equilibria under the explicit assumption that the eigenvalue λ_{N-1} is*

[125] *Stability Regions of Nonlinear Dynamical Systems*. Hsiao-Dong Chiang et al. (2015).
[126] "Multiequilibria Analysis for a Class of Collective Decision-Making". Angela Fontan et al. (2018).

simple. On the other hand, our proof relies on a topological argument concerning stability boundaries and does not require this simplicity assumption, thus holding for all network topologies such that $\lambda_{N-1} > 0$, including those with repeated eigenvalues.

This final proposition shows that the condition for synchronization is in a sense, necessary. If the network is not sufficiently connected (i.e., λ_{N-1} is not large and negative) and the signal function is sufficiently nonlinear (i.e., K is large), disagreement is not just possible, but inevitable by construction. This leads directly to the formulation of a sharp threshold.

4.5 A Sharp Threshold for Global Synchronization

The preceding analysis has prepared us for the central question of this chapter. We have established that disagreement can only emerge if the signal function s is sufficiently nonlinear to possess unstable fixed points (Theorem 4.4.1). Conversely, we have shown that certain network topologies with strong connectivity properties can enforce consensus regardless of the signal function's shape (e.g., Proposition 4.4.4, Lemma 4.4.5). Finally, we demonstrated that for any non-trivial graph, there *exists* a signal function that can induce non-synchronized equilibria (Proposition 4.4.6).

These findings frame a tension between the cohesive influence of the network topology and the divisive potential of the signal function's nonlinearity. The analysis of necessary conditions in Theorem 4.4.1 suggests that the *shape* of s is central, while the constructive proof in Proposition 4.4.6 points to a trade-off between network connectivity and the Lipschitz constant K.

This leads to the ultimate question for full synchronization: Do these conditions collapse into a single, sharp threshold? Specifically, is the condition $K\lambda_{N-1} < 1$ sufficient to *guarantee* full synchronization for *any* admissible signal function, thereby overpowering the specific shape properties previously found to be necessary for disagreement?

This section provides the definitive answer. We will prove that the condition $K\lambda_{N-1} < 1$ is indeed sufficient to confine all system equilibria to the synchronized manifold, regardless of the specific shape of the signal function s. This result, when combined with Proposition 4.4.6, formally establishes the condition as a sharp threshold. In doing so, we generalize the condition found in [126, Theorem 4], demonstrating that a guarantee previously established for a restrictive class of smooth, sigmoidal interactions holds for the broad class of all admissible interaction laws considered in this work. The result is a condition that elegantly predicts whether a given system will globally synchronize or permit structured disagreement, as illustrated in the simulations of Figure 4.4.

Theorem 4.5.1 (Sharp Threshold for Synchronization)**.** *Let K be the Lipschitz constant of s. If $K\lambda_{N-1} < 1$, then equilibria of* (COSA) *are FSE.*

Proof. The proof proceeds by contraposition. We assume the existence of an NFSE, denoted by $\boldsymbol{x}^*$, and we will demonstrate that this assumption necessarily implies $K\lambda_{N-1} \geq 1$.

An equilibrium point $\boldsymbol{x}^*$ of the system (COSA) must satisfy the condition (3.9). i.e., $\boldsymbol{x}^* = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}^*)$. Since $\boldsymbol{x}^*$ is an NFSE, by Theorem 4.4.1-(ii), there exists $c^\circ \in \mathrm{Fix}^\circ(s)$ such that the components of $\boldsymbol{x}^*$ and $\boldsymbol{s}(\boldsymbol{x}^*)$ are not all on one side of c°. Specifically, the sets $\mathcal{I} = \{i \mid s(x_i^*) < c^\circ\}$ and $\mathcal{J} = \{j \mid s(x_j^*) > c^\circ\}$ are non-empty.

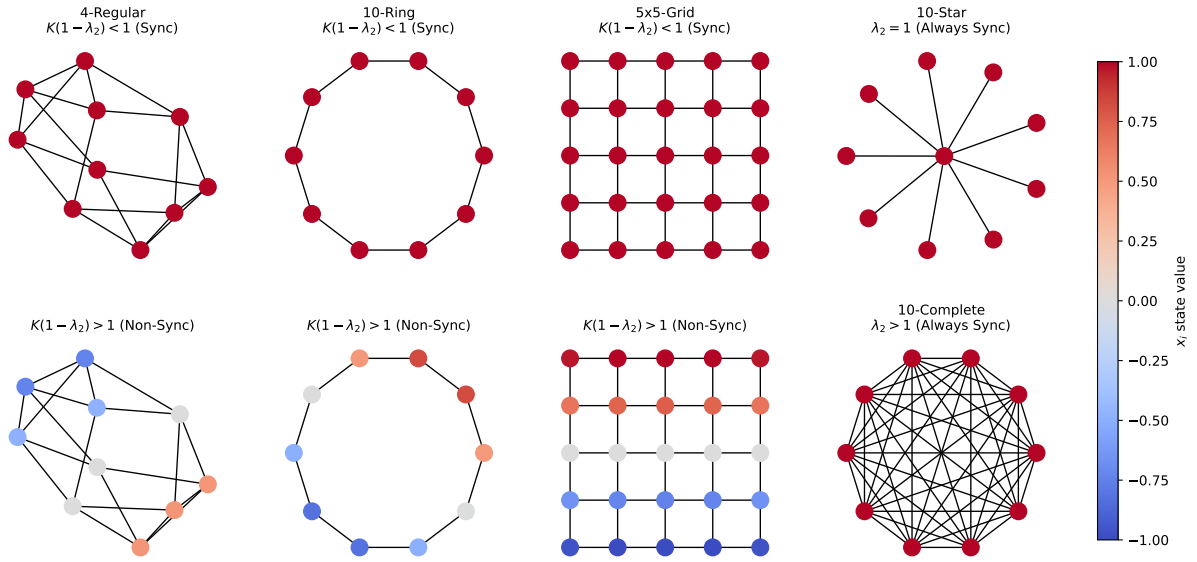


Figure 4.4: Asymptotic behavior of (COSA) on various graph topologies under different conditions for the synchronization parameter $K(1-\lambda_2)$. Node colors represent the agent state $x_i \in [-1, 1]$, as indicated by the color bar. For each graph topology in columns 1-3, the top panel shows the final state when $K(1-\lambda_2) < 1$, resulting in global synchronization (Theorem 4.5.1). The bottom panel shows the final state for the same graph topology when $K(1-\lambda_2) > 1$, using a signal function $s(x)$ that facilitates non-synchronized equilibria (Theorem 4.4.1), leading to persistent disagreement patterns. **(Column 1)** 4-Regular graph ($N = 10$, $\lambda_2 \approx 0.42$): (Top) $K_1 = 2.6$. (Bottom) $K_2 = 2.7$. **(Column 2)** 10-Ring graph ($N = 10$, $\lambda_2 \approx 0.19$): (Top) $K_1 = 1.1$. (Bottom) $K_2 = 1.3$. **(Column 3)** 5x5-Grid graph ($N = 25$, $\lambda_2 \approx 0.11$): (Top) $K_1 = 1.1$. (Bottom) $K_2 = 1.4$. **(Column 4)** Graphs with strong synchronizing properties: (Top) 10-Star graph ($N = 10$, $\lambda_2 = 1$), always synchronizes. (Bottom) 10-Complete graph ($N = 10$, $\lambda_2 = N/(N-1) > 1$), always synchronizes due to high connectivity. Both are shown for parameters $K = 50$ still ensuring non fully synchronized equilibria. The signal function for both top and bottom panels is $s(x) = \max(-1, \min(1, Kx))$. Initial condition are the same for both top and bottom panels and chosen to ensure non fully synchronized equilibria, and all simulations run until reaching a steady state.

We now center the equilibrium equation around this fixed point c°. Let $\boldsymbol{v}^* = \boldsymbol{x}^* - c^\circ \mathbf{1}$. Subtracting $c^\circ \mathbf{1}$ from the equilibrium equation and using $\boldsymbol{D}^{-1}\boldsymbol{A}\mathbf{1} = \mathbf{1}$ yields:

$$\boldsymbol{v}^* = \boldsymbol{D}^{-1}\boldsymbol{A}(\boldsymbol{s}(\boldsymbol{x}^*) - c^\circ \mathbf{1}).$$

To recover $\boldsymbol{v}^*$ in the right hand side, we define a diagonal matrix $\boldsymbol{M} = \mathrm{diag}(m_1, \ldots, m_N)$, where

$$m_i = \begin{cases} \dfrac{s(x_i^*) - c^\circ}{x_i^* - c^\circ} & \text{if } x_i^* \neq c^\circ, \\ K & \text{if } x_i^* = c^\circ. \end{cases}$$

The Lipschitz continuity of s ensures that each diagonal entry satisfies $0 \leq m_i \leq K$. By construction, we have $\boldsymbol{s}(\boldsymbol{x}^*) - c^\circ \mathbf{1} = \boldsymbol{M}\boldsymbol{v}^*$, which transforms the centered equation into:

$$\boldsymbol{v}^* = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{M}\boldsymbol{v}^*. \tag{4.8}$$

To proceed, we define the vector $\boldsymbol{y}^* = \boldsymbol{M}^{1/2}\boldsymbol{v}^*$ and show that it must be non-zero and contain both positive and negative entries. We argue by contradiction. Suppose $\boldsymbol{y}^*$ contains no positive entries, i.e., $y_i^* \leq 0$ for all $i \in \mathcal{V}$. This implies that for any index $j \in \mathcal{J}$ (where $v_j^* > 0$), the corresponding entry m_j must be zero. By definition of m_j, this would mean $s(x_j^*) = c^\circ$. This contradicts the fact that $\mathcal{J}$ is the set of indices where $s(x_j^*) > c^\circ$. Therefore, our assumption is false, and $\boldsymbol{y}^*$ must contain at least one positive entry. A symmetric argument using the set $\mathcal{I}$ shows it must also contain at least one negative entry. Consequently, $\boldsymbol{y}^*$ is a non-zero vector with mixed-sign entries.

Left-multiplying (4.8) by $\boldsymbol{M}^{1/2}$ yields $\boldsymbol{y}^* = \boldsymbol{M}^{1/2}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{M}^{1/2}\boldsymbol{y}^*$. This reveals that 1 is an eigenvalue of the matrix $\boldsymbol{\Theta} = \boldsymbol{M}^{1/2}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{M}^{1/2}$, with corresponding eigenvector $\boldsymbol{y}^*$.

To analyze the eigenvalues of $\boldsymbol{\Theta}$, we relate it to a symmetric matrix. Let $\tilde{\boldsymbol{A}} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ be the symmetrically normalized adjacency matrix, which is symmetric and has the same eigenvalues as $\boldsymbol{D}^{-1}\boldsymbol{A}$ by similarity. The matrix $\boldsymbol{\Theta}$ is similar to the symmetric matrix $\tilde{\boldsymbol{\Theta}} = \boldsymbol{M}^{1/2}\tilde{\boldsymbol{A}}\boldsymbol{M}^{1/2}$ via the transformation $\tilde{\boldsymbol{\Theta}} = (\boldsymbol{D}^{1/2}\boldsymbol{M}^{-1/2})\boldsymbol{\Theta}(\boldsymbol{M}^{1/2}\boldsymbol{D}^{-1/2})$. Therefore, $\boldsymbol{\Theta}$ and $\tilde{\boldsymbol{\Theta}}$ share the same real eigenvalues, which means 1 is also an eigenvalue of the symmetric matrix $\tilde{\boldsymbol{\Theta}}$. The corresponding eigenvector is $\boldsymbol{z}^* = \boldsymbol{D}^{1/2}\boldsymbol{y}^*$. Since $\boldsymbol{D}^{1/2}$ has strictly positive diagonal entries, $\boldsymbol{z}^*$ inherits the mixed-sign property of $\boldsymbol{v}^*$.

Since $\tilde{\boldsymbol{\Theta}}$ is a nonnegative matrix, by Perron-Frobenius Theorem, the eigenvectors of $\tilde{\boldsymbol{\Theta}}$ associated with the largest eigenvalue $\lambda_N(\tilde{\boldsymbol{\Theta}})$ must be nonnegative. Since $\boldsymbol{z}^*$ has mixed-sign, $\boldsymbol{z}^*$ must be associated with $\lambda_i(\tilde{\boldsymbol{\Theta}})$ for $i \in \mathcal{V}$ such that $\lambda_i(\tilde{\boldsymbol{\Theta}}) \neq \lambda_N(\tilde{\boldsymbol{\Theta}})$. This implies that $1 < \lambda_N(\tilde{\boldsymbol{\Theta}})$, and therefore 1 must be one of the lower eigenvalues, so $1 \leq \lambda_{N-1}(\tilde{\boldsymbol{\Theta}})$.

Finally, we use Ostrowski's inequality for matrix congruence on $\tilde{\boldsymbol{\Theta}} = (\boldsymbol{M}^{1/2})^\top\tilde{\boldsymbol{A}}(\boldsymbol{M}^{1/2})$. This theorem guarantees that $\lambda_k(\tilde{\boldsymbol{\Theta}}) = \theta_k\lambda_k(\tilde{\boldsymbol{A}})$ for some scalar $\theta_k \in [\min_i(m_i), \max_i(m_i)]$. Since $0 \leq m_i \leq K$ for all i, we have $\theta_k \in [0, K]$. Applying this to the $(N-1)$-th eigenvalue, and noting that $\lambda_{N-1}(\tilde{\boldsymbol{A}}) = \lambda_{N-1}$, we get:

$$1 \leq \lambda_{N-1}(\tilde{\boldsymbol{\Theta}}) = \theta_{N-1}\lambda_{N-1}(\tilde{\boldsymbol{A}}) \leq K\lambda_{N-1}.$$

The existence of an NFSE has led to the necessary condition $K\lambda_{N-1} \geq 1$. By contraposition, if $K\lambda_{N-1} < 1$, no such equilibria can exist. □

Remark 4.5.1 (Generalization of Necessity Condition)**.** *This theorem and its proof generalize the necessity condition to obtain NFSE with sigmoidal signals in [126, Theorem 4] in two key aspects. First, our result applies to a much broader class of signal functions, requiring only that s be non-decreasing and Lipschitz-continuous, whereas the model in [126] assumes smooth, odd, and sigmoidal nonlinearities. Second, our proof technique correctly establishes the non-strict inequality $K\lambda_{N-1} \geq 1$ as the necessary condition for disagreement, refining the strict inequality derived for their specific model.*

A direct consequence of this theorem is that, when the threshold condition is met, the synchronized manifold becomes the global attractor for the system's dynamics.

Corollary 4.5.2 (Global Attractivity of the FSM)**.** *Let K be the Lipschitz constant of s such that $K\lambda_{N-1} < 1$. Then, for almost all initial conditions $\boldsymbol{x}(0) \in \mathcal{X}$, one has $\lim_{t\to\infty} \boldsymbol{x}(t) \in \mathcal{S}$.*

If additionally, s is strictly increasing, this is verified for all initial conditions in $\mathcal{X}$. Meaning that the FSM $\mathcal{S}$ is globally attractive.

Proof. From Proposition 3.3.2, we have that for almost all initial conditions, the associated trajectory converges to an equilibrium. Since the only equilibria are synchronized equilibria, the fully synchronized manifold $\mathcal{S}$ is globally attractive. If s is strictly increasing and smooth,

from Proposition 3.3.4, we have the previous verified for all initial conditions, leading to the global attractiveness of $\mathcal{S}$. □

Remark 4.5.2 (Attractivity vs. Lyapunov Stability)**.** *It is mandatory to distinguish the global attractivity of $\mathcal{S}$ guaranteed by Corollary 4.5.2 from its Lyapunov stability. Even when the threshold is met, it is possible for trajectories to transiently move away from $\mathcal{S}$ before eventually returning, which precludes a general statement on global asymptotic stability.*

4.6 Global exponential stability of the FSM

While the preceding results guarantee the global attractivity of the synchronized manifold $\mathcal{S}$ under the condition $K\lambda_{N-1} < 1$, this provides no assurance on the transient behavior of the system. As noted in Remark 4.5.2, trajectories may exhibit significant excursions from the manifold before eventually converging.

We now establish a stricter condition for this stronger property: global exponential stability. The following theorem demonstrates that if the cohesive strength of the network, quantified by the full spectrum of its disagreement modes, is sufficient to overcome the maximal divisive influence of the signal function, then not only will the system synchronize, but any disagreement will decay exponentially fast.

Theorem 4.6.1 (Exponential stability of the FSM)**.** *Let K be the Lipschitz constant of s and $\alpha := K \max_{i\in\{1,\dots,N-1\}} |\lambda_i|$. If $\alpha < 1$, then the fully synchronized manifold $\mathcal{S}$ is globally exponentially stable.*

Proof. Let us work the frame of the synchronization error vector $\boldsymbol{e} = \boldsymbol{x} - \bar{x}\mathbf{1}$ defined in Definition 4.2.3. When the error is zero, the agents are synchronized.

Consider the Lyapunov function defined in (4.4) given by $V(\boldsymbol{x}) = \boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{e}/2$ for which $V(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x} = \bar{x}\mathbf{1}$ (i.e. $\boldsymbol{e} = 0$). The time derivative of the Lyapunov function is given by

$$\dot{V}(\boldsymbol{x}) = \boldsymbol{e}^\top \boldsymbol{D}\dot{\boldsymbol{e}} = (\boldsymbol{x} - \bar{x}\mathbf{1})^\top \boldsymbol{D}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}\right),$$

note that by Lemma 4.2.4, one has $\boldsymbol{e}^\top \boldsymbol{A}\mathbf{1} = \boldsymbol{e}^\top \boldsymbol{D}\mathbf{1} = 0$. Then, by substituting $\boldsymbol{x} = \boldsymbol{e} + \bar{x}\mathbf{1}$ and $\boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{s}(\boldsymbol{x}) - s(\bar{x})\mathbf{1} + s(\bar{x})\mathbf{1}$ in the dynamics of the error, one has

$$\begin{aligned}\dot{V}(\boldsymbol{x}) &= \boldsymbol{e}^\top \boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{D}\left(\boldsymbol{x} - \bar{x}\mathbf{1} + \bar{x}\mathbf{1}\right) = \boldsymbol{e}^\top \boldsymbol{A}\left(\boldsymbol{s}(\boldsymbol{x}) - s(\bar{x})\mathbf{1} + s(\bar{x})\mathbf{1}\right) - \boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{e} \\ &= \boldsymbol{e}^\top \boldsymbol{A}\left(\boldsymbol{s}(\boldsymbol{x}) - s(\bar{x})\mathbf{1}\right) - \boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{e}.\end{aligned}$$

The Lipschitz condition of s implies $|s(x_i) - s(\bar{x})| \leq K|x_i - \bar{x}| = K|e_i|$, for all $i \in \{1, \dots, N\}$. This allows us to write:

$$\boldsymbol{s}(\boldsymbol{x}) - s(\bar{x})\mathbf{1} = K\boldsymbol{\Theta}\boldsymbol{e}$$

where $\boldsymbol{\Theta} = \operatorname{diag}(\theta_1, \dots, \theta_N)$ is a diagonal matrix with state-dependent entries satisfying $\theta_i = (s(x_i) - s(\bar{x}))/K(x_i - \bar{x}) = (s(x_i) - s(\bar{x}))/Ke_i$ for $e_i \neq 0$ and $\theta_i = 1$ for $e_i = 0$. Note that $0 \leq \theta_i \leq 1$.

Substituting this into the expression for $\dot{V}$ yields:

$$\dot{V}(\boldsymbol{x}) = K\boldsymbol{e}^\top \boldsymbol{A}\boldsymbol{\Theta}\boldsymbol{e} - \boldsymbol{e}^\top \boldsymbol{D}\boldsymbol{e}$$

Let us perform a change of coordinates, $\tilde{\boldsymbol{e}} = \boldsymbol{D}^{1/2}\boldsymbol{e}$. The expression for $\dot{V}$ becomes:

$$\dot{V}(\boldsymbol{x}) = K\tilde{\boldsymbol{e}}^\top \tilde{\boldsymbol{A}}\boldsymbol{\Theta}\tilde{\boldsymbol{e}} - \|\tilde{\boldsymbol{e}}\|_2^2$$

where $\tilde{\boldsymbol{A}} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ is the normalized adjacency matrix. As in proof of Theorem 4.5.1, $\tilde{\boldsymbol{A}}$ is similar to $\boldsymbol{D}^{-1}\boldsymbol{A}$, therefore, they share the same eigenvalues $\lambda_1, \dots, \lambda_N$.

In the transformed coordinates, the normalized eigenvector of $\tilde{\boldsymbol{A}}$ for $\lambda_N = 1$ is $\boldsymbol{v}_N = \boldsymbol{D}^{1/2}\mathbf{1}$. The transformed error vector $\tilde{\boldsymbol{e}}$ is orthogonal to $\boldsymbol{v}_N$:

$$\tilde{\boldsymbol{e}}^\top \boldsymbol{v}_N = (\boldsymbol{D}^{1/2}\boldsymbol{x})^\top \boldsymbol{D}^{1/2}\mathbf{1} = \boldsymbol{x}^\top \boldsymbol{D}\mathbf{1} = 0.$$

Thus, $\tilde{\boldsymbol{e}}$ lies in the subspace spanned by the eigenvectors $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_{N-1}\}$ of $\tilde{\boldsymbol{A}}$, which are orthogonal to $\boldsymbol{v}_N$.

We now bound the quadratic form by applying the Cauchy-Schwarz inequality:

$$\tilde{\boldsymbol{e}}^\top \tilde{\boldsymbol{A}}\boldsymbol{\Theta}\tilde{\boldsymbol{e}} = (\tilde{\boldsymbol{A}}\tilde{\boldsymbol{e}})^\top(\boldsymbol{\Theta}\tilde{\boldsymbol{e}}) \leq \|\tilde{\boldsymbol{A}}\tilde{\boldsymbol{e}}\|_2 \|\boldsymbol{\Theta}\tilde{\boldsymbol{e}}\|_2.$$

Let us bound the two norms separately. Since $\boldsymbol{\Theta}$ is a diagonal matrix with entries $|\theta_i| \leq 1$, its induced 2-norm satisfies $\|\boldsymbol{\Theta}\|_2 = \max_i |\theta_i| \leq 1$. Thus, $\|\boldsymbol{\Theta}\tilde{\boldsymbol{e}}\|_2 \leq \|\tilde{\boldsymbol{e}}\|_2$.

For $\|\tilde{\boldsymbol{A}}\tilde{\boldsymbol{e}}\|_2$, we use the spectral decomposition of $\tilde{\boldsymbol{A}}$ since it is symmetric. Let $\{\boldsymbol{v}_i\}_{i=1}^N$ be an orthonormal basis of eigenvectors for $\tilde{\boldsymbol{A}}$. As established, $\tilde{\boldsymbol{e}}$ has no component along $\boldsymbol{v}_N$, so it can be written as $\tilde{\boldsymbol{e}} = \sum_{i=1}^{N-1} c_i \boldsymbol{v}_i$, where $c_i = \tilde{\boldsymbol{e}}^\top \boldsymbol{v}_i$. Then,

$$\|\tilde{\boldsymbol{A}}\tilde{\boldsymbol{e}}\|_2 = \Big\| \sum_{i=1}^{N-1} c_i \lambda_i \boldsymbol{v}_i \Big\|_2 \leq \max_{j\in\{1,\dots,N-1\}} |\lambda_j| \|\tilde{\boldsymbol{e}}\|_2$$

Combining these two bounds, we get:

$$\tilde{\boldsymbol{x}}^\top \tilde{\boldsymbol{A}}\boldsymbol{\Theta}\tilde{\boldsymbol{x}} \leq \Big(\max_{i\in\{1,\dots,N-1\}} |\lambda_i| \Big) \|\tilde{\boldsymbol{x}}\|_2^2.$$

Substituting this back into the expression for $\dot{V}(\boldsymbol{e})$:

$$\dot{V}(\boldsymbol{x}) \leq K\Big(\max_{i\in\{1,\dots,N-1\}} |\lambda_i| \Big) \|\tilde{\boldsymbol{e}}\|_2^2 - \|\tilde{\boldsymbol{x}}\|_2^2 = 2\Big(K \max_{i\in\{1,\dots,N-1\}} |\lambda_i| - 1 \Big) V(\boldsymbol{e}) \coloneqq -2\alpha V(\boldsymbol{x})$$

since $\|\tilde{\boldsymbol{e}}\|_2^2 = \|\boldsymbol{e}\|_{\boldsymbol{D}}^2 = 2V(\boldsymbol{x})$ and $\alpha = K \max_{i\in\{1,\dots,N-1\}} |\lambda_i| - 1 > 0$ by the condition of the theorem. By Grönwall's inequality, we have:

$$V(\boldsymbol{x}(t)) \leq V(\boldsymbol{x}(0)) e^{-2\alpha t} \Leftrightarrow \|\boldsymbol{e}(t)\|_{\boldsymbol{D}} \leq \|\boldsymbol{e}(0)\|_{\boldsymbol{D}} e^{-\alpha t}.$$

This shows that $V(\boldsymbol{x}(t)) \to 0$ exponentially, which implies $\|\boldsymbol{e}(t)\|_{\boldsymbol{D}} \to 0$ exponentially. By Proposition 4.2.2, this implies that $\boldsymbol{x}(t) \to \mathcal{S}$ exponentially. Thus, the fully synchronized manifold $\mathcal{S}$ is globally exponentially stable. □

This theorem provides an analytical condition for robust synchronization. Unlike the underestimation case, which guarantees consensus for any connected graph, this condition explicitly depends on the graph's spectral properties through $\max_{i<N} |\lambda_i|$. It highlights that even for signal functions that are not globally contractive and may even be locally expansive (e.g., overestimation, as shown in Figure 4.5), robust global synchronization is still achievable, provided the network's connectivity is sufficiently strong to dominate the nonlinearity of the agent interactions. This result thus unifies the roles of network structure and agent behavior in determining the system's ultimate convergence properties.

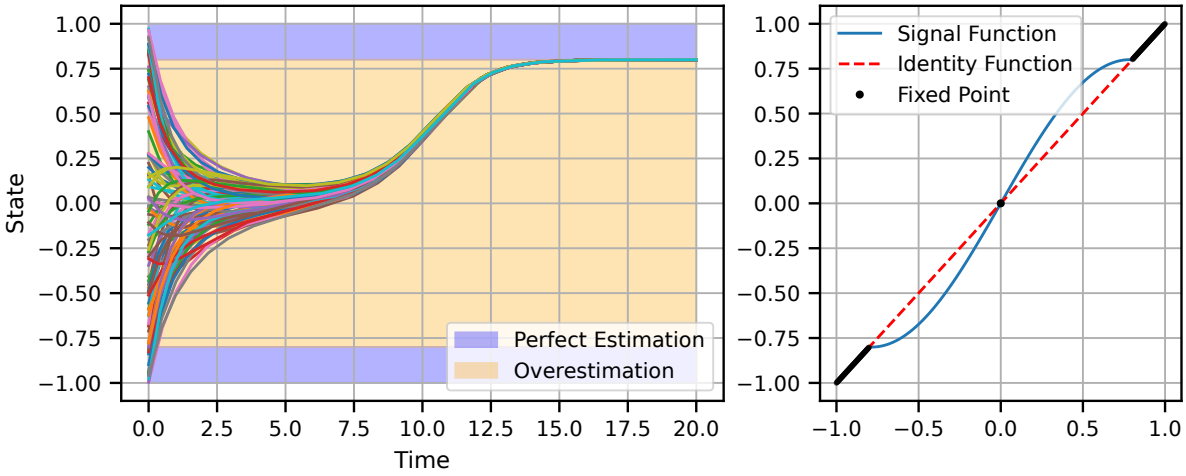


Figure 4.5: State evolution for agents with a globally overestimating signal under the same conditions as in Figure 4.3, $s(x) = 0.8\sin(2x)/\sin(1.6)$ for $x \in [-0.8, 0.8]$ and $s(x) = x$ otherwise.

4.7 Concluding Remarks on Full Synchronization

This chapter has provided a comprehensive and rigorous analysis of full synchronization in the (COSA) framework. By moving from local stability analysis to global convergence guarantees, we have delimited the precise conditions under which a network of agents is assured to reach a state of complete consensus.

The analysis was founded upon the introduction of a dynamically-motivated geometry for the state space, defined by the degree-weighted inner product. This framework, which renders the system's core operator self-adjoint, proved essential for a tractable Lyapunov-based analysis of the synchronization error. Our first cornerstone result demonstrated that the stability of a FSE in the full N-dimensional space is completely determined by the stability of its corresponding fixed point in the one-dimensional scalar dynamics (FS) studied in Section 3.6.1 of the previous chapter. This reduction principle allowed us to lift the simple, intuitive properties of the signal function s to characterize the complex, high-dimensional system behavior. We established that the basins of attraction for FSEs are defined by invariant hypercubes whose boundaries are formed by the system's unstable FSEs.

This led to the chapter's central investigation: the trade-off between the forces of the network topology and the divisive nonlinearities of the signal function. The analysis reached a sharp threshold for global synchronization, in the condition $K\lambda_{N-1} < 1$. This inequality elegantly captures the balance between the network's algebraic connectivity, represented by λ_{N-1}, and the signal function's maximal nonlinearity, quantified by the Lipschitz constant K. When the network's cohesive influence dominates, disagreement equilibria are excluded, and the fully synchronized manifold $\mathcal{S}$ becomes the global attractor for the system. Furthermore, we provided a stricter, verifiable condition, $K \max_{i<N} |\lambda_i| < 1$, that guarantees not just attractivity but global exponential stability of the FSM.

Having characterized the conditions for convergence to the synchronized manifold, the logical next step is to investigate the alternative: disagreement. The strict threshold that guarantees synchronization also implies its failure: when $K\lambda_{N-1} \geq 1$, the emergence of disagreement becomes possible. The subsequent chapter will explore this other side of the dichotomy, analyzing

the structure and stability of the NFSE that arise when the system's dynamics get away from the consensus manifold. The tools developed here, particularly the understanding of unstable fixed points as the seeds of complex behavior, will prove indispensable in uncovering the rich patterns of structured disagreement and emergent clustering in the network.

Chapter 5

Partial Synchronization in COSA

5.1 Introduction

The preceding chapter established the conditions for the system (COSA) to reach full synchronization. We now investigate the alternative: the emergence of Non-Fully Synchronized Equilibria (NFSE), where agents settle into states of persistent disagreement. These states are often highly structured, with agents forming coherent subgroups. This phenomenon of *local synchronization*, or clustering, is a form of self-organization that occurs when global consensus is not reached.

This chapter examines two forms of local synchronization, each requiring a different analytical approach:

- **Perfect Local Synchronization.** This is a strict form where agents within a given cluster converge to the *exact* same state. We will show that this is not a general behavior but is instead imposed by specific symmetries in the network's structure. We will use the theory of *equitable partitions* to formally characterize these required symmetries and show how they allow for a simplified analysis of the system dynamics.
- **Approximated Local Synchronization.** This is a more general and robust form of clustering where agents within a group do not perfectly agree, but their disagreement remains small and bounded. To analyze this, we will model clusters as subsystems influenced by the rest of the network. We will then use the framework of *Input-to-State Stability (ISS)* to formally quantify how a cluster's internal cohesion is robust to these external influences.

By analyzing both perfect and approximated local synchronization, this chapter provides a detailed understanding of the system's behavior beyond global consensus. The analysis begins with the stricter case of perfect synchronization, which is tied to network structure, before moving to the more general framework of robust, approximate clustering. Together, these results explain how complex patterns of agreement and disagreement emerge and persist in nonlinear multi-agent systems.

5.2 Perfect Local Synchronization: The Constraint of Structural Symmetry

When the system settles into a Non-Fully Synchronized Equilibrium (NFSE), the disagreement is often structured. This section investigates the strictest form of this structure: **Perfect Local Synchronization**, where agents within specific subgroups agree exactly.

We establish that this phenomenon is not a generic outcome but is instead strictly constrained by symmetries in the network topology. Our central result is that the existence of such "exact" clusters, which we term *Cell-Synchronized Equilibria (CSE)*, requires the underlying graph to possess an *regular partition* condition that is achieved by an *equitable partition*. This constraint is significant because it reduces the search for these equilibria to a combinatorial problem dependent only on the graph's structure, independent of the specific signal function s.

5.2.1 From NFSE to Cell-Synchronized Equilibria (CSE)

We formally define the specific subset of NFSEs that exhibit perfect internal cohesion. This requires the concept of a partition, as defined in the preliminaries (Chapter 2.2.4). We recall the definition of a partition here for convenience.

Definition 5.2.1 (Partition). *A **partition** of the set $\mathcal{V}$ is a collection of non-empty subsets $\pi = \{\mathcal{C}_1, \dots, \mathcal{C}_M\}$, called **cells**, such that $\bigcup_{k=1}^{M} \mathcal{C}_k = \mathcal{V}$ and $\mathcal{C}_k \cap \mathcal{C}_l = \emptyset$ for all $k \neq l$. We denote the number of agents in cell $\mathcal{C}_k$ as its cardinality, $N_k = |\mathcal{C}_k|$.*

Our interest lies in the states that respect such a partition. We formally define the concept of a *Cell-Synchronized (CS)* state and a *Cell-Synchronized Equilibrium (CSE)*.

Definition 5.2.2 (Cell-Synchronized (CS) State). *A state vector $\boldsymbol{x} \in \mathcal{X}$ is said to be **Cell-Synchronized (CS)** with respect to a partition π if for every cell $\mathcal{C}_k \in \pi$, all agents within that cell have the same state value. That is, $x_i = x_j$ for all $i, j \in \mathcal{C}_k$.*

Moreover, we define the cell-synchronized vector $\boldsymbol{y} = (y_1, \dots, y_M)^\top \in \mathcal{Y} := [-1, 1]^M$ such that $x_i = y_k$ for all $i \in \mathcal{C}_k$ and all $k \in \{1, \dots, M\}$.

Definition 5.2.3 (Cell-Synchronized Equilibrium (CSE)). *A equilibrium point $\boldsymbol{x}^* \in \mathcal{X}$ of the dynamics (COSA) is said to be a **Cell-Synchronized Equilibrium (CSE)** with respect to a partition π if it is a cell-synchronized state with respect to π.*

Moreover, we define the vector $\boldsymbol{y}^* = (y_1^*, \dots, y_M^*)^\top \in \mathcal{Y}$ such that $x_i^* = y_k^*$ for all $i \in \mathcal{C}_k$ and all $k \in \{1, \dots, M\}$.

Remark 5.2.1. *The CSE are special cases of the NFSE. They are much more restrictive since they require the states of all agents in a cell to be exactly the same, assumption that is not necessarily true for the NFSE in general (except for the trivial partition where each cell is a singleton).*

A CSE represents a configuration where the network naturally segregates into M distinct, internally synchronized clusters. The critical research question guiding this section is: *What exact structural property must a partition π satisfy to ensure that a CSE based on π is a possible equilibrium outcome for **any** non-trivial signal function s?*

5.2.2 The Necessity of Structural Symmetry for CSEs

We seek a condition on the communication graph $\mathcal{G}$ that is necessary and sufficient for the existence of CSEs, independent of the specific functional form of s.

Let us consider a partition $\pi = \{\mathcal{C}_1, \dots, \mathcal{C}_M\}$ of the vertex set $\mathcal{V}$. Our objective is to find the conditions under which the system (COSA) can support a CSE $\boldsymbol{x}^*$. Such an equilibrium is characterized by the existence of a vector of cell values $\boldsymbol{y}^* = (y_1^*, \dots, y_M^*)^\top \in \mathcal{Y}$ such that for any cell $\mathcal{C}_k$, the equilibrium state of every agent $i \in \mathcal{C}_k$ is identical, i.e., $x_i^* = y_k^*$.

The equilibrium condition for any agent $i \in \mathcal{V}$ is $\dot{x}_i = 0$, which from (3.10) yields:

$$x_i^* = \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} s(x_j^*).$$

Let us apply our cell synchronization assumption to this equation for an arbitrary agent $i \in \mathcal{C}_k$. The left-hand side becomes $x_i^* = y_k^*$. The sum on the right-hand side can be reorganized by grouping the neighbors of i according to the cells they belong to. Let b_{il} denote the number of neighbors of agent i that are in cell $\mathcal{C}_l$, i.e., $b_{il} = |\mathcal{N}_i \cap \mathcal{C}_l|$. The equilibrium condition for agent $i \in \mathcal{C}_k$ can thus be written as:

$$y_k^* = \frac{1}{d_i} \sum_{l=1}^{M} b_{il} s(y_l^*). \tag{5.1}$$

The core of our premise is that this condition must hold for *every* agent within the cell $\mathcal{C}_k$. If we consider another agent $i' \in \mathcal{C}_k$ (where $i' \neq i$), its equilibrium state must also be y_k^*, leading to a similar constraint:

$$y_k^* = \frac{1}{d_{i'}} \sum_{l=1}^{M} b_{i'l} s(y_l^*). \tag{5.2}$$

For equations (5.1) and (5.2) to be simultaneously satisfied for any arbitrary vector of cell values $\boldsymbol{y}^*$ (and thus for any signal s), the following identity must hold:

$$\sum_{l=1}^{M} \left(\frac{b_{il}}{d_i} - \frac{b_{i'l}}{d_{i'}} \right) s(y_l^*) = 0, \quad \forall \boldsymbol{y}^* \in \mathcal{Y}.$$

For this linear combination of the functions $s(y_l^*)$ to be identically zero for any non-trivial choice of s, the coefficients must be zero. This gives us the necessary condition:

$$\frac{b_{il}}{d_i} = \frac{b_{i'l}}{d_{i'}}, \quad \forall l \in \{1, \dots, M\} \text{ and } \forall i, i' \in \mathcal{C}_k. \tag{5.3}$$

This condition defines a *regular partition* (also known as *lumpable partition* in Markov chains theory). However, our goal is to link the dynamics to the purely combinatorial and structural properties of the graph. A more natural condition is to require the degree of neighbors in the cell to be constant. Imposing this condition, one obtains the definition of an **equitable partition** (Definition 2.2.24). We recall the definition of an equitable partition here for convenience.

Definition 5.2.4 (Equitable Partition (EP))**.** *A partition $\pi = \{\mathcal{C}_1, \dots, \mathcal{C}_M\}$ of $\mathcal{V}$ is called **equitable** if for any two cells $\mathcal{C}_k$ and $\mathcal{C}_l$, the number of neighbors in $\mathcal{C}_l$ of a vertex $i \in \mathcal{C}_k$ is a constant b_{kl} that is independent of the specific choice of $i \in \mathcal{C}_k$. Moreover, an equitable partition is* connected *if each cell $\mathcal{C}_k$ induces a connected subgraph $\mathcal{G}_k$.*

Remark 5.2.2 (Equitable Partition Implies Regular Partition)**.** *The property of being an equitable partition is a stronger, combinatorial constraint that directly implies the algebraic necessity condition (5.3).*

If π is equitable, then $b_{il} = b_{i'l} = b_{kl}$. Furthermore, the total degree of any agent in $\mathcal{C}_k$ is constant, $d_i = d_{i'} = d_{\mathcal{C}_k} = \sum_{l=1}^{M} b_{kl}$. Consequently, the ratios simplify:

$$\frac{b_{il}}{d_i} = \frac{b_{kl}}{d_{\mathcal{C}_k}} = \frac{b_{i'l}}{d_{i'}}.$$

Thus, the equitable partition is a sufficient structural prerequisite for a network to support CSEs independent of the specific signal function s.



(a) Closed Twins. (b) Open Twins.

Figure 5.1: Illustration of structural twins.

This derivation establishes a link between the system's dynamics and its underlying structure. The necessary condition for a CSE to exist independently of the signal function s is that the agents form a *regular partition*. As we have shown, this algebraic condition is guaranteed by the stronger, purely combinatorial property of an *equitable partition*.

Therefore, the existence of an equitable partition is the key structural property that enables perfect local synchronization. This insight is important because it transforms the search for exact clusters from a complex analytical problem into a more tractable combinatorial search for specific symmetries within the graph. This structural constraint is not just a restriction; it is what allows for a complete reduction of the system's dynamics, as we will demonstrate next.

5.2.3 A Foundational Example: Twin Node Synchronization

To build intuition for perfect local synchronization, we begin with the simplest non-trivial example of a cell in an EP: a pair of *twin* nodes. This concept provides a direct link between local network symmetry and the resulting dynamics.

Definition 5.2.5 (Twin Nodes)**.** *Two distinct agents $i, j \in \mathcal{V}$ are called **twins**, denoted $i \sim j$, if their neighborhoods, excluding each other, are identical: $\mathcal{N}_i \setminus \{j\} = \mathcal{N}_j \setminus \{i\}$.*

On Figure 5.1, we illustrate the concept of twins. The left figure shows a closed twin, where the two nodes have the same neighborhood except for each other. The right figure shows an open twin, where the two nodes have the same neighborhood except for each other. We treat both cases similarly.

Example 5.2.1 (Star Graph)**.** *For the star graph, every nodes except the center node are twins.*

This definition identifies a local symmetry in the network. An immediate consequence is that twin nodes must have the same degree, i.e., if $i \sim j$, then $d_i = d_j$. Based on this equivalence relation, we define the twin synchronized manifold.

Definition 5.2.6 (Twin Synchronized Manifold)**.** *The **twin synchronized manifold** $\mathcal{S}_\sim$ is the set of all states $\boldsymbol{x} \in \mathcal{X}$ such that $x_i = x_j$ for all $i, j \in \mathcal{V}$ such that $i \sim j$. That is,*

$$\mathcal{S}_\sim = \{\boldsymbol{x} \in \mathcal{X} \mid x_i = x_j,\ \forall i, j \in \mathcal{V} \text{ such that } i \sim j\}\,.$$

Remark 5.2.3. *If two nodes i and j are twins, then the partition containing the cell $\mathcal{C}_k = \{i, j\}$ (with all other nodes in singleton cells) is equitable.*

The manifold $\mathcal{S}_\sim$ consists of all states in which twin agents have identical state values. This elementary symmetry has a direct dynamic consequence.

Proposition 5.2.7. *The set $\mathcal{S}_\sim$ is forward invariant for the dynamics* (COSA)*. i.e., if $\boldsymbol{x}(0) \in \mathcal{S}_\sim$, then $\boldsymbol{x}(t) \in \mathcal{S}_\sim$ for all $t \geq 0$.*

Proof. Assume $\boldsymbol{x} \in \mathcal{S}_\sim$ and consider a pair of twin agents $i, j \in \mathcal{V}$ (i.e., $i \sim j$) with $x_i = x_j$. Since $\mathcal{G}$ is undirected and $i \sim j$, it follows that the degrees of these agents are equal ($d_i = d_j$). Consider the difference in their dynamics:

$$\begin{aligned}\dot{x}_i - \dot{x}_j &= \frac{1}{d_i}\sum_{k\in\mathcal{N}_i} s(x_k) - x_i - \frac{1}{d_j}\sum_{k\in\mathcal{N}_j} s(x_k) + x_j \\ &= \frac{1}{d_i}\Big[\sum_{k\in\mathcal{N}_i\setminus\{j\}} s(x_k) - \sum_{k\in\mathcal{N}_j\setminus\{i\}} s(x_k) - s(x_i) + s(x_j)\Big] = 0,\end{aligned}$$

since $\mathcal{N}_i \setminus \{j\} = \mathcal{N}_j \setminus \{i\}$ and for $\boldsymbol{x} \in \mathcal{S}_\sim$ we have $x_i = x_j$. Thus, if $x_i(0) = x_j(0)$, it follows that $x_i(t) = x_j(t)$ for all $t \geq 0$, which establishes the forward invariance of $\mathcal{S}_\sim$. □

Forward invariance ensures that synchronized twins remain synchronized (Proposition 5.2.7). A stronger result shows that even twins who start with different states will asymptotically synchronize, meaning the twin synchronized manifold is not just invariant but also attractive (Proposition 5.2.8).

Proposition 5.2.8 (Twin Synchronization)**.** *Let $i, j \in \mathcal{V}$ be twin agents. Then, under the dynamics* (COSA)*, they asymptotically synchronize, i.e., $\lim_{t\to\infty}(x_i(t) - x_j(t)) = 0$. Moreover, the convergence is exponential.*

Proof. Consider the difference $\delta_{ij} = x_i - x_j$. Since i and j are twins, they have the same degree, $d_i = d_j$. The dynamics of the difference are given by:

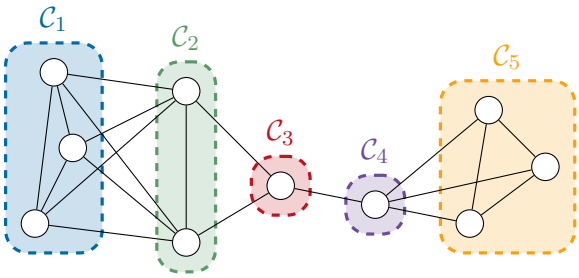
$$\begin{aligned}\dot{\delta}_{ij} = \dot{x}_i - \dot{x}_j &= \frac{1}{d_i}\sum_{k\in\mathcal{N}_i} s(x_k) - \frac{1}{d_j}\sum_{k\in\mathcal{N}_j} s(x_k) + x_j - x_i \\ &= \frac{1}{d_i}\Big[\sum_{k\in\mathcal{N}_i\setminus\{j\}} s(x_k) - \sum_{k\in\mathcal{N}_j\setminus\{i\}} s(x_k) - s(x_i) + s(x_j)\Big] - \delta_{ij} = \frac{1}{d_i}\left[s(x_j) - s(x_i)\right] - \delta_{ij}.\end{aligned}$$

Consider the Lyapunov function $V(\delta_{ij}) = \delta_{ij}^2/2$. Its time derivative is

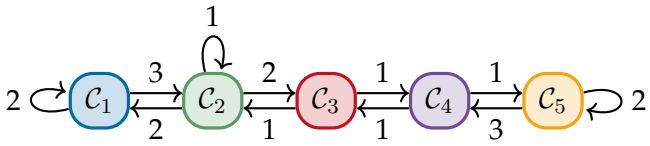
$$\dot{V}(\delta_{ij}) = \delta_{ij}\dot{\delta}_{ij} = \delta_{ij}\left[\frac{1}{d_i}[s(x_j) - s(x_i)] - \delta_{ij}\right] = -\frac{1}{d_i}(x_i - x_j)(s(x_i) - s(x_j)) - \delta_{ij}^2.$$

Since s is non-decreasing, the term $(x_i - x_j)(s(x_i) - s(x_j)) \geq 0$. Therefore, $\dot{V}(\delta_{ij}) \leq -\delta_{ij}^2 = -2V(\delta_{ij})$, which guarantees exponential convergence of $\delta_{ij}(t)$ to zero as $t \to \infty$. □

This result confirms that the local symmetry inherent in twin nodes enforces local synchronization regardless of the signal function's nonlinearity, provided only that s is non-decreasing. This principle generalizes from twin pairs to all cells of an equitable partition, a property that is central to the analysis in the next section.



(a) A graph with a connected equitable partition.



(b) Quotient graph of the graph (a) with the partition $\pi = \{\mathcal{C}_1, \ldots, \mathcal{C}_5\}$.

Figure 5.2: Example of a graph with equitable partition and its quotient graph.

5.3 Analysis of Perfect Clusters via the Quotient Graph

Having established in Section 5.2 that an EP is a sufficient structural requirement for exact clustering, we now demonstrate its analytical consequence: the EP structure permits an exact reduction of the system's dimensionality. This reduction allows us to map the complex, N-dimensional dynamics of clustered agents onto a simpler, M-dimensional *quotient system*, where each node represents an entire cell of agents. The analysis of this tractable, lower-dimensional system provides a complete characterization of the existence, multiplicity, and stability of all possible CSEs.

5.3.1 Model Reduction: The Quotient Dynamics

The model reduction strategy relies on a key property: the subspace of states synchronized according to an equitable partition is an invariant manifold for the dynamics.

Definition 5.3.1 (Cell Synchronized Manifold)**.** *Given a graph $\mathcal{G}$ with an EP π, the **Cell Synchronized Manifold**, $\mathcal{S}_\pi$, is the subspace of states that are cell-synchronized with respect to π:*

$$\mathcal{S}_\pi = \{\boldsymbol{x} \in \mathcal{X} \mid \boldsymbol{x} = \boldsymbol{P}\boldsymbol{y} \text{ for some } \boldsymbol{y} \in \mathcal{Y}\},$$

where $\boldsymbol{P}$ is the characteristic matrix of π.

Proposition 5.3.2 (Forward Invariance of $\mathcal{S}_\pi$)**.** *Let π be an EP of the graph $\mathcal{G}$. The manifold $\mathcal{S}_\pi$ is forward invariant with respect to the dynamics* (COSA).

Proof. Consider an initial state $\boldsymbol{x} \in \mathcal{S}_\pi$. By definition, there exists a vector $\boldsymbol{y} \in \mathcal{Y}$ such that $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{y}$. This means that for any cell $\mathcal{C}_k$, all agents $i \in \mathcal{C}_k$ have the same initial state, $x_i = y_k$.

To prove invariance, we must show that the time derivatives $\dot{x}_i$ are identical for all agents i within the same cell $\mathcal{C}_k$ at any time t where $\boldsymbol{x} \in \mathcal{S}_\pi$. Let us examine the dynamics for an

arbitrary agent $i \in \mathcal{C}_k$:

$$\dot{x}_i = \frac{1}{d_i} \sum_{j=1}^{N} a_{ij} s(x_j) - x_i.$$

Since $\boldsymbol{x} \in \mathcal{S}_\pi$, we can substitute $x_j = y_l$ for all $j \in \mathcal{C}_l$, and $x_i = y_k$. We can group the sum over neighbors by the cells they belong to:

$$\sum_{j=1}^{N} a_{ij} s(x_j) = \sum_{l=1}^{M} \sum_{j \in \mathcal{N}_i \cap \mathcal{C}_l} s(y_l) = \sum_{l=1}^{M} b_{kl} s(y_l).$$

The last step follows from Definition 5.2.4, as the number of neighbors of $i \in \mathcal{C}_k$ in cell $\mathcal{C}_l$ is the constant b_{kl}. As established in the Remark in Section 3, the degree d_i is also synchronized for all $i \in \mathcal{C}_k$, equal to $d_{\mathcal{C}_k}$. Substituting these back, we get:

$$\dot{x}_i = \frac{1}{d_{\mathcal{C}_k}} \sum_{l=1}^{M} b_{kl} s(y_l) - y_k.$$

This expression for $\dot{x}_i$ depends only on the index of its cell, k, and not on the specific agent i. Therefore, for any two agents $i, i' \in \mathcal{C}_k$, we have $\dot{x}_i = \dot{x}_{i'}$. Since they share the same initial state and the same derivative, their states will remain identical for all times. This holds for all cells, so $\boldsymbol{x}$ remains in $\mathcal{S}_\pi$. □

The invariance of $\mathcal{S}_\pi$ allows us to project the high-dimensional system onto the dynamics of the cell values $\boldsymbol{y}(t)$, yielding a lower-dimensional and analytically tractable **quotient system**.

Proposition 5.3.3 (Quotient Dynamics)**.** *Let π be an EP of the graph $\mathcal{G}$ with characteristic matrix $\boldsymbol{P}$. For any state $\boldsymbol{x} \in \mathcal{S}_\pi$ with $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{y}$, the dynamics of the cell values $\boldsymbol{y}$ are governed by the M-dimensional system:*

$$\dot{\boldsymbol{y}} = \tilde{\boldsymbol{D}}^{-1} \boldsymbol{B} \boldsymbol{s}(\boldsymbol{y}) - \boldsymbol{y}, \tag{5.4}$$

where $\boldsymbol{B}$ is the adjacency matrix of the quotient graph and $\tilde{\boldsymbol{D}} = \mathrm{diag}(d_{\mathcal{C}_1}, \ldots, d_{\mathcal{C}_M})$ is the diagonal matrix of cell degrees.

Proof. We start with the collective dynamics (COSA): $\dot{\boldsymbol{x}} = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}$. Since $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{y}$, we have $\dot{\boldsymbol{x}} = \boldsymbol{P}\dot{\boldsymbol{y}}$. The signal function acts element-wise, so for any $i \in \mathcal{C}_k$, we have $s(x_i) = s(y_k)$. This implies the vector relation $\boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{s}(\boldsymbol{P}\boldsymbol{y}) = \boldsymbol{P}\boldsymbol{s}(\boldsymbol{y})$. Substituting these into the dynamics yields:

$$\boldsymbol{P}\dot{\boldsymbol{y}} = \boldsymbol{D}^{-1}\boldsymbol{A}(\boldsymbol{P}\boldsymbol{s}(\boldsymbol{y})) - \boldsymbol{P}\boldsymbol{y}.$$

Now we apply the identities from Lemma 2.2.26. First, we replace $\boldsymbol{A}\boldsymbol{P}$ with $\boldsymbol{P}\boldsymbol{B}$:

$$\boldsymbol{P}\dot{\boldsymbol{y}} = \boldsymbol{D}^{-1}\boldsymbol{P}\boldsymbol{B}\boldsymbol{s}(\boldsymbol{y}) - \boldsymbol{P}\boldsymbol{y}.$$

Next, we replace $\boldsymbol{D}^{-1}\boldsymbol{P}$ with $\boldsymbol{P}\tilde{\boldsymbol{D}}^{-1}$:

$$\boldsymbol{P}\dot{\boldsymbol{y}} = \boldsymbol{P}\tilde{\boldsymbol{D}}^{-1}\boldsymbol{B}\boldsymbol{s}(\boldsymbol{y}) - \boldsymbol{P}\boldsymbol{y} = \boldsymbol{P}\left(\tilde{\boldsymbol{D}}^{-1}\boldsymbol{B}\boldsymbol{s}(\boldsymbol{y}) - \boldsymbol{y}\right).$$

The columns of the characteristic matrix $\boldsymbol{P}$ are linearly independent. We can therefore left-multiply by the pseudo-inverse $\boldsymbol{P}^\dagger = (\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{P}^\top$ to isolate $\dot{\boldsymbol{y}}$, which yields the quotient dynamics (5.4). □

Remark 5.3.1 (Equilibrium Equivalence)**.** *This reduction to the quotient system establishes an equivalence: a cell-synchronized state $\boldsymbol{x}^* = \boldsymbol{P}\boldsymbol{y}^*$ is a CSE of the N-dimensional system if and only if its cell-value vector $\boldsymbol{y}^*$ is an equilibrium of the M-dimensional quotient system. This result is the building block of our analysis, as it reduces the problem of finding and analyzing complex clustered equilibria in large networks to studying the equilibria of a much smaller, more tractable dynamical system.*

5.3.2 Analysis of the Quotient System

Having derived the quotient dynamics in (5.4), we can now use this tractable, lower-dimensional model to fully characterize the CSEs.

Existence of Local Agreement Equilibria

Before characterizing the properties of these equilibria, we first establish their guaranteed existence. The following proposition confirms that the quotient system always possesses at least one equilibrium, which ensures that at least one CSE exists for any equitable partition π.

Proposition 5.3.4. *The quotient dynamics (5.4) possess at least one equilibrium point $\boldsymbol{y}^* \in \mathcal{Y}$.*

Proof. The function $\boldsymbol{g}(\boldsymbol{y}) = \tilde{\boldsymbol{D}}^{-1}\boldsymbol{B}\boldsymbol{s}(\boldsymbol{y})$ is continuous on the compact, convex set $\mathcal{Y}$ since s is continuous. Furthermore, coordinates of $\boldsymbol{g}(\boldsymbol{y})$ are convex combinations of values $s(y_j) \in [-1, 1]$, thus $\boldsymbol{g}(\boldsymbol{y}) \in \mathcal{Y}$. By Brouwer's Fixed Point Theorem, $\boldsymbol{g}$ must have at least one fixed point $\boldsymbol{y}^* \in \mathcal{Y}$. □

Characterization and Bifurcation of Equilibria

While existence is guaranteed, the nonlinear nature of s can lead to the emergence of multiple CSE. To understand this phenomenon, we analyze the equilibrium condition for a single cell, which reveals a trade-off between internal feedback and external influence.

The equilibrium condition $\boldsymbol{y}^* = \tilde{\boldsymbol{D}}^{-1}\boldsymbol{B}\boldsymbol{s}(\boldsymbol{y}^*)$ can be written for a single cell $\mathcal{C}_k$ as:

$$y_k^* = \frac{1}{d_{\mathcal{C}_k}} \sum_{l=1}^{M} b_{kl} s(y_l^*) = \frac{b_{kk}}{d_{\mathcal{C}_k}} s(y_k^*) + \sum_{l \neq k} \frac{b_{kl}}{d_{\mathcal{C}_k}} s(y_l^*).$$

This expression motivates the definition of two key quantities for each cell $\mathcal{C}_k$:

– The *internal influence ratio*,

$$\mu_k = b_{kk}/d_{\mathcal{C}_k},$$

which is the fraction of a cell's connections that are internal to itself.

– The *external influence*,

$$m_k(\boldsymbol{y}_{-k}) = \frac{1}{d_{\mathcal{C}_k} - b_{kk}} \sum_{l \neq k} b_{kl} s(y_l),$$

which represents the weighted average of signals received from all other cells. Note that $d_{\mathcal{C}_k} - b_{kk} = \sum_{l \neq k} b_{kl}$.

With these definitions, the equilibrium condition for cell $\mathcal{C}_k$ can be rewritten as a balance equation:

$$y_k^* = \mu_k s(y_k^*) + (1 - \mu_k) m_k(\boldsymbol{y}_{-k}^*). \tag{5.5}$$

For a fixed external influence m_k, finding the equilibrium y_k^* is equivalent to finding a fixed point of the one-dimensional function $g_k(y_k) = \mu_k s(y_k) + (1-\mu_k)m_k$. The following lemma provides the condition for this fixed point to be unique.

Lemma 5.3.5. *Let $\mu \in [0,1)$ and $m \in [-1,1]$. The function $g(x) = \mu s(x) + (1-\mu)m$ has a unique fixed point in $[-1,1]$ for every $m \in [-1,1]$ if and only if the function $h(x) = \mu s(x) - x$ is strictly decreasing on $[-1,1]$.*

Proof. A fixed point x^* satisfies $x^* = \mu s(x^*) + (1-\mu)m$. Define $f(x) = x - \mu s(x)$, so the condition is $f(x^*) = C$, where $C = (1-\mu)m$. As m varies in $[-1,1]$, C varies in the interval $I = [-(1-\mu), 1-\mu]$. The lemma is equivalent to stating: the equation $f(x) = C$ has a unique solution $x \in [-1,1]$ for every $C \in I$ if and only if $f(x)$ is strictly increasing on $[-1,1]$ (since $f = -h$).
($\Leftarrow$) Assume $f(x)$ is strictly increasing. Then f is injective, guaranteeing at most one solution for any C. We evaluate the endpoints: $f(-1) = -1 - \mu s(-1) \leq -1 + \mu = -(1-\mu)$ and $f(1) = 1 - \mu s(1) \geq 1 - \mu$. Since f is continuous, by the Intermediate Value Theorem, the range $f([-1,1])$ contains $[f(-1), f(1)]$, which includes I. Thus, for every $C \in I$, a solution exists. Existence plus injectivity yields uniqueness.
($\Rightarrow$) We do a proof by contradiction and assume $f(x)$ is not strictly increasing on $[-1,1]$. Since f is continuous, if it is not strictly increasing, it is not injective. Therefore, there exist $x_1 < x_2$ in $[-1,1]$ such that $f(x_1) = f(x_2)$. Let $C_0 = f(x_1) = f(x_2)$. If $C_0 \in I = [-(1-\mu), 1-\mu]$, then for the value $m_0 = C_0/(1-\mu) \in [-1,1]$, the equation $f(x) = C_0$ (and thus the fixed point equation for $g_{m_0}(x)$) has at least two solutions (x_1 and x_2), contradicting the premise that the fixed point is unique for all $m \in [-1,1]$. Since any continuous, non-strictly increasing function connecting $f(-1) \leq \min(I)$ to $f(1) \geq \max(I)$ must necessarily attain some value $C_0 \in I$ at least twice, the premise implies that f must be strictly increasing.
Therefore, the condition holds if and only if $f(x)$ is strictly increasing, which is equivalent to $h(x) = -f(x)$ being strictly decreasing. □

This lemma allows us to connect the network structure (via μ_k) to the potential for multiple equilibria. A high internal influence μ_k combined with a strong nonlinearity in s can cause $h(x)$ to lose its monotonicity, leading to bifurcations.

Proposition 5.3.6. *Let $\mathcal{G}$ be a graph with an EP $\pi = \{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$ and K be the minimum Lipschitz constant of s. Then, the following hold true.*
- *If $\mu_k K < 1$, then $g_k(y_k, \boldsymbol{y}_{-k})$ admits a unique fixed point for any fixed $\boldsymbol{y}_{-k} \in [-1,1]^{M-1}$.*
- *If $\mu_k K > 1$, there exists $\boldsymbol{y}_{-k} \in [-1,1]^{M-1}$ such that $g_k(y_k, \boldsymbol{y}_{-k})$ admits multiple fixed points in its first argument y_k.*

Proof. Let us first assume that $\mu_k K < 1$. One can verify that g_k is a contraction mapping on its first argument y_k for any fixed $\boldsymbol{y}_{-k} \in [-1,1]^{M-1}$. Indeed, for any $y_k, \tilde{y}_k \in [-1,1]$,

$$|g_k(y_k, \boldsymbol{y}_{-k}) - g_k(\tilde{y}_k, \boldsymbol{y}_{-k})| = |\mu_k(s(y_k) - s(\tilde{y}_k))| \leq \mu_k K \, |y_k - \tilde{y}_k| < |y_k - \tilde{y}_k| \, .$$

Therefore, g_k admits a unique fixed point in its first argument y_k for any $\boldsymbol{y}_{-k} \in [-1,1]^{M-1}$.

Assume now that $\mu_k > 1/K$. Then, $1/\mu_k \leq K$ and since s is non-decreasing and with minimum Lipschitz constant K, we have $\sup_{y_a \neq y_b}(s(y_b) - s(y_a))/(y_b - y_a) = K$. This means

that there exist $y_a, y_b \in [-1, 1]$ with $y_a < y_b$ such that $(s(y_b) - s(y_a))/(y_b - y_a) \geq K \geq 1/\mu_k$. Then, by rearranging,

$$\frac{s(y_b) - s(y_a)}{y_b - y_a} \geq \frac{1}{\mu_k} \Leftrightarrow \mu_k s(y_b) - y_b \geq \mu_k s(y_a) - y_a \Leftrightarrow h(y_b) \geq h(y_a),$$

where $h(y) = \mu_k s(y) - y$. Thus, since $m_k(\boldsymbol{y}_{-k}) \in [-1, 1]$, if such $y_a < y_b$ exist, we have $h(y_b) \geq h(y_a)$ implying that $h(y_k)$ is not strictly decreasing. Then, from Lemma 5.3.5, g_k admits multiple fixed points in its first argument y_k for some values of $\boldsymbol{y}_{-k} \in [-1, 1]^{M-1}$. □

Remark 5.3.2 (The Role of Internal Cohesion). *Proposition 5.3.6 provides a key insight: cells with strong internal cohesion (large μ_k) are more prone to bistability and complex behaviors for the whole system. When internal feedback is strong enough to overcome the maximal amplification of the signal function (K), a cell can sustain its own state even against conflicting external signals. This is the core mechanism for symmetry breaking and the emergence of distinct, stable clusters for* (COSA).

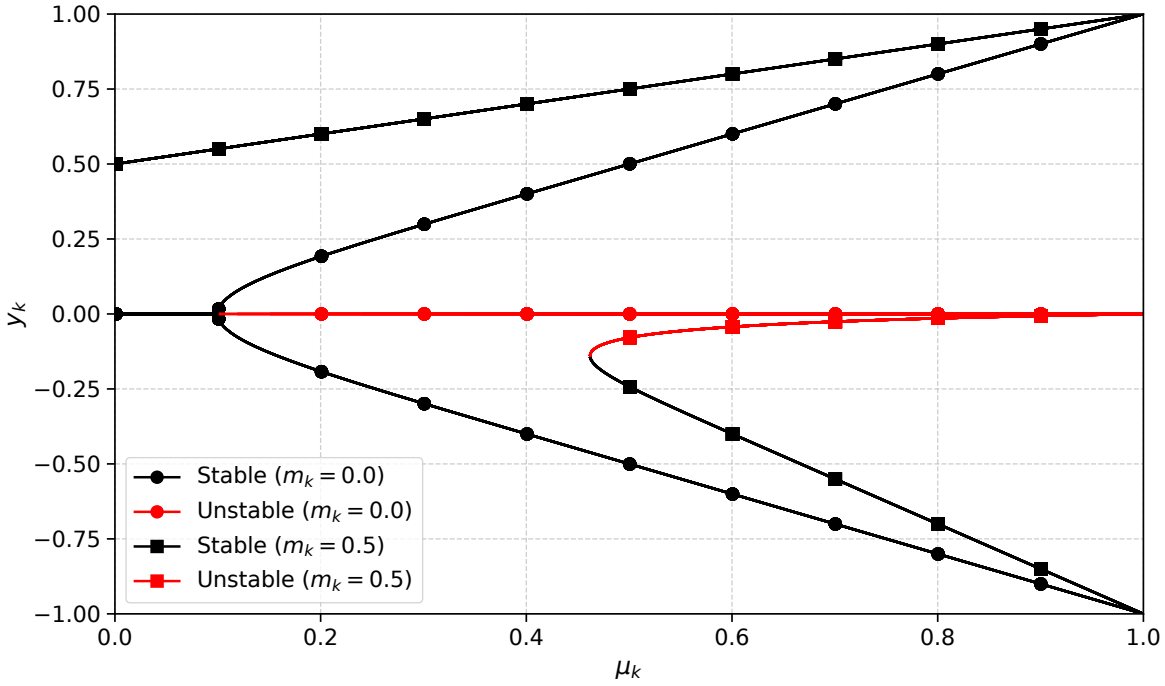


Figure 5.3: Bifurcation diagram of (5.5) for different values of $m(\boldsymbol{y}_{-k})$ with $s(x) = \tanh(10x)$. The red dashed line represents unstable fixed points, and the black line represents stable fixed points. The markers represent the fixed points for the different values of m. For $m = 0$, there is a pitchfork bifurcation at $\mu_k = 0.1$ while for $m = 0.5$, there is a perturbed pitchfork bifurcation at $\mu_k = 0.5$.

5.3.3 Stability of Local Agreement Equilibria

The analysis in Section 5.3.2 guaranteed the existence of at least one CSE. However, the existence of such an equilibrium does not guarantee its stability. In this section, we derive the conditions for the local exponential stability of a CSE, $\boldsymbol{x}^* = \boldsymbol{P}\boldsymbol{y}^*$ **under the additional assumption that**

the signal function s is continuously differentiable in a neighborhood of each component of $\boldsymbol{y}^*$. The core of our analysis is to decompose the stability problem, leveraging the invariant manifold structure of $\mathcal{S}_\pi$. Although the local stability of any equilibrium $\boldsymbol{x}^*$ can be studied by looking at the Jacobian matrix of the full system, this may not be suitable or computationally feasible for large networks. On the other hand, the quotient system is low-dimensional and thus easier to handle.

The structure of a CSE on the manifold $\mathcal{S}_\pi$ means that perturbations can be decomposed into two distinct types, each corresponding to a different stability requirement.

Quotient Stability

First, we analyze the system's response to perturbations *along* the manifold $\mathcal{S}_\pi$, where the average state of each cluster is perturbed but internal synchronization is preserved. A perturbation of this type moves the state from $\boldsymbol{x}^* = \boldsymbol{P}\boldsymbol{y}^*$ to a nearby state $\boldsymbol{x} = \boldsymbol{P}(\boldsymbol{y}^* + \boldsymbol{\varepsilon})$.

The stability against these perturbations is entirely determined by the dynamics of $\boldsymbol{y}^*$ as an equilibrium of the M-dimensional quotient dynamics (5.4). Consequently, for the cluster configuration to be stable, the equilibrium $\boldsymbol{y}^*$ must be a stable equilibrium of the quotient dynamics. To assess this, we linearize the system around the equilibrium $\boldsymbol{y}^*$, which yields the *quotient Jacobian matrix*:

$$\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*) := \tilde{\boldsymbol{D}}^{-1}\boldsymbol{B}\boldsymbol{s}'(\boldsymbol{y}^*) - \boldsymbol{I}_M, \tag{5.6}$$

where $\boldsymbol{s}'(\boldsymbol{y}^*) = \operatorname{diag}(s'(y_1^*), \dots, s'(y_M^*))$.

We establish the relationship between the quotient and full system spectra via the following lemma.

Lemma 5.3.7 (Spectral Inclusion). *Let $\boldsymbol{x}^* = \boldsymbol{P}\boldsymbol{y}^*$ be a CSE. The spectrum of the quotient Jacobian $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$ is a subset of the spectrum of the full Jacobian $\boldsymbol{J}(\boldsymbol{x}^*)$. Specifically, for every eigenpair $(\lambda, \boldsymbol{v})$ of $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$, $(\lambda, \boldsymbol{P}\boldsymbol{v})$ is an eigenpair of $\boldsymbol{J}(\boldsymbol{x}^*)$.*

Proof. We show that the Jacobian matrices satisfy the intertwining relation $\boldsymbol{J}(\boldsymbol{x}^*)\boldsymbol{P} = \boldsymbol{P}\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$. Recall the definitions $\boldsymbol{J}(\boldsymbol{x}^*) = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}'(\boldsymbol{x}^*) - \boldsymbol{I}_N$ and $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*) = \tilde{\boldsymbol{D}}^{-1}\boldsymbol{B}\boldsymbol{s}'(\boldsymbol{y}^*) - \boldsymbol{I}_M$.

Since $\boldsymbol{x}^* \in \mathcal{S}_\pi$, we have $\boldsymbol{s}'(\boldsymbol{x}^*)\boldsymbol{P} = \boldsymbol{P}\boldsymbol{s}'(\boldsymbol{y}^*)$. Using this and the identities from Lemma 2.2.26:

$$\begin{aligned}
\boldsymbol{J}(\boldsymbol{x}^*)\boldsymbol{P} &= (\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}'(\boldsymbol{x}^*) - \boldsymbol{I}_N)\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{s}'(\boldsymbol{y}^*) - \boldsymbol{P} \\
&= \boldsymbol{D}^{-1}\boldsymbol{P}\boldsymbol{B}\boldsymbol{s}'(\boldsymbol{y}^*) - \boldsymbol{P} \quad (\text{using } \boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{B}) \\
&= \boldsymbol{P}\tilde{\boldsymbol{D}}^{-1}\boldsymbol{B}\boldsymbol{s}'(\boldsymbol{y}^*) - \boldsymbol{P} \quad (\text{using } \boldsymbol{D}^{-1}\boldsymbol{P} = \boldsymbol{P}\tilde{\boldsymbol{D}}^{-1}) \\
&= \boldsymbol{P}\left(\tilde{\boldsymbol{D}}^{-1}\boldsymbol{B}\boldsymbol{s}'(\boldsymbol{y}^*) - \boldsymbol{I}_M\right) \\
&= \boldsymbol{P}\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*).
\end{aligned}$$

Let $(\lambda, \boldsymbol{v})$ be an eigenpair of $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$. Multiplying the relation by $\boldsymbol{v}$ from the right yields:

$$\boldsymbol{J}(\boldsymbol{x}^*)(\boldsymbol{P}\boldsymbol{v}) = \boldsymbol{P}\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)\boldsymbol{v} = \boldsymbol{P}(\lambda\boldsymbol{v}) = \lambda(\boldsymbol{P}\boldsymbol{v}).$$

Since $\boldsymbol{P}$ has full column rank, $\boldsymbol{P}\boldsymbol{v} \neq \boldsymbol{0}$, proving that $\boldsymbol{P}\boldsymbol{v}$ is an eigenvector of $\boldsymbol{J}(\boldsymbol{x}^*)$ associated with λ. □

This inclusion allows us to formally state the necessary and sufficient condition for stability with respect to perturbations within the manifold $\mathcal{S}_\pi$.

Proposition 5.3.8 (Quotient Stability). *Let $\boldsymbol{x}^* = \boldsymbol{P}\boldsymbol{y}^*$ be a CSE for an EP π. Assume that the signal function s is continuously differentiable in a neighborhood of each component of $\boldsymbol{y}^*$. The equilibrium $\boldsymbol{x}^*$ is exponentially stable with respect to perturbations restricted to the invariant subspace $\mathcal{S}_\pi$ if and only if the quotient Jacobian $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$ is Hurwitz.*

Proof. The dynamics restricted to the invariant subspace $\mathcal{S}_\pi$ is topologically conjugate to the linear system $\dot{\boldsymbol{z}} = \boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)\boldsymbol{z}$ via the coordinate transformation $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{z}$. Therefore, the restricted system is exponentially stable if and only if all eigenvalues of $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$ have strictly negative real parts (i.e., $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$ is Hurwitz). This is a direct consequence of the linearization principle [117]. □

Transversal Stability

Second, we must analyze perturbations *transversal* to the manifold $\mathcal{S}_\pi$, which break the synchronization within one or more cells. For example, two agents in the same cell are perturbed from their common state such that their new states are no longer identical. Transversal stability requires that the system dynamics actively suppress such internal disagreements, forcing agents within any given cell to re-synchronize. If this condition is not met, a cell could "break apart" and lose its coherence, even if the cell's average state remains stable according to the quotient dynamics. These properties fall into two categories:

– **Internal Cohesion:** Each cell $\mathcal{C}_k$ must possess internal dynamics that are inherently stable.

– **Inter-Cell Coupling:** Disagreements are coupled across cells. A perturbation in cell $\mathcal{C}_j$ can propagate and destabilize cell $\mathcal{C}_k$.

To emphasize this decomposition, let us consider a small perturbation $\boldsymbol{\delta}$ around $\boldsymbol{x}^*$. Its linearized dynamics are $\dot{\boldsymbol{\delta}} = \boldsymbol{J}(\boldsymbol{x}^*)\boldsymbol{\delta}$. We focus on the block of equations corresponding to the agents in a cell $\mathcal{C}_k$. The perturbation sub-vector $\boldsymbol{\delta}_k$ evolves according to:

$$\dot{\boldsymbol{\delta}}_k = (\boldsymbol{J}(\boldsymbol{x}^*)\boldsymbol{\delta})_k = \sum_{j=1}^{M} \boldsymbol{J}_{kj}\boldsymbol{\delta}_j = \boldsymbol{J}_{kk}\boldsymbol{\delta}_k + \sum_{j\neq k} \boldsymbol{J}_{kj}\boldsymbol{\delta}_j,$$

where $\boldsymbol{J}_{kj}$ is the (k,j)-th block of the Jacobian $\boldsymbol{J}(\boldsymbol{x}^*)$. From the definition of $\boldsymbol{J}(\boldsymbol{x}^*)$, these blocks are $\boldsymbol{J}_{kk} = \frac{s'(y_k^*)}{d_{\mathcal{C}_k}}\boldsymbol{A}_k - \boldsymbol{I}_{N_k}$ and $\boldsymbol{J}_{kj} = \frac{s'(y_j^*)}{d_{\mathcal{C}_k}}\boldsymbol{A}_{kj}$ for $k \neq j$.

Moreover, the dynamics for the disagreement vector $\boldsymbol{e}_k = \boldsymbol{\Pi}_k^\perp \boldsymbol{\delta}_k$ is found by applying the projection operator $\boldsymbol{\Pi}_k^\perp := \boldsymbol{I}_{N_k} - \frac{1}{N_k}\mathbf{1}_{N_k}\mathbf{1}_{N_k}^\top$ to the dynamics of $\boldsymbol{\delta}_k$:

$$\dot{\boldsymbol{e}}_k = \boldsymbol{\Pi}_k^\perp \dot{\boldsymbol{\delta}}_k = \underbrace{\boldsymbol{\Pi}_k^\perp \left(\frac{s'(y_k^*)}{d_{\mathcal{C}_k}}\boldsymbol{A}_k - \boldsymbol{I}_{N_k} \right) \boldsymbol{\delta}_k}_{\text{Internal}} + \underbrace{\sum_{j\neq k} \boldsymbol{\Pi}_k^\perp \frac{s'(y_j^*)}{d_{\mathcal{C}_k}}\boldsymbol{A}_{kj}\boldsymbol{\delta}_j}_{\text{Inter-Cell}}. \tag{5.7}$$

Internal Cohesion Stability In this part, we focus only on perturbations restricted to a specific cell $\mathcal{C}_k$ for some $k \in \{1, \dots, M\}$. This means that for all $j \neq k$, $\boldsymbol{\delta}_j = \mathbf{0}$. To measure the internal cohesion, we define the *internal error decay rate* of the cell k,

$$\alpha_k := 1 - \frac{s'(y_k^*)}{d_{\mathcal{C}_k}}\lambda_{N_k-1}(\boldsymbol{A}_k), \tag{5.8}$$

[117] *Nonlinear Systems*. Hassan K Khalil. (2002).

where $\lambda_{N_k-1}(\boldsymbol{A}_k)$ is the second largest eigenvalue of the induced subgraph adjacency matrix $\boldsymbol{A}_k$.

Proposition 5.3.9 (Internal Cohesion). *Let $\boldsymbol{x}^* = \boldsymbol{P}\boldsymbol{y}^*$ be a CSE resulting from a connected EP π. Then, the equilibrium $\boldsymbol{x}^*$ is locally exponentially stable for any perturbation restricted to cell $\mathcal{C}_k$ if and only if $\alpha_k > 0$.*

Proof. Considering a perturbation only inside the cell of a CSE is equivalent to studying the error dynamics (5.7) without the inter-cell part, since $\boldsymbol{\delta}_j = \mathbf{0}$ for all $j \neq k$.

The EP structure implies that the adjacency matrix of the induced subgraph of $\mathcal{C}_k$, $\boldsymbol{A}_k$, has a constant row sum b_{kk}. This means $\boldsymbol{A}_k \mathbf{1}_{N_k} = b_{kk}\mathbf{1}_{N_k}$, so $\boldsymbol{A}_k$ and the projector $\boldsymbol{\Pi}_k^{\perp}$ commute. Therefore:

$$\boldsymbol{\Pi}_k^{\perp}\boldsymbol{J}_{kk}\boldsymbol{\delta}_k = \boldsymbol{\Pi}_k^{\perp}\left(\frac{s'(y_k^*)}{d_{\mathcal{C}_k}}\boldsymbol{A}_k - \boldsymbol{I}_{N_k}\right)\boldsymbol{\delta}_k = \left(\frac{s'(y_k^*)}{d_{\mathcal{C}_k}}\boldsymbol{A}_k - \boldsymbol{I}_{N_k}\right)\boldsymbol{\Pi}_k^{\perp}\boldsymbol{\delta}_k = \left(\frac{s'(y_k^*)}{d_{\mathcal{C}_k}}\boldsymbol{A}_k - \boldsymbol{I}_{N_k}\right)\boldsymbol{e}_k.$$

Thus, the internal error dynamics of cell k around $\boldsymbol{x}^*$ is given by

$$\dot{\boldsymbol{e}}_k = \left(\frac{s'(y_k^*)}{d_{\mathcal{C}_k}}\boldsymbol{A}_k - \boldsymbol{I}_{N_k}\right)\boldsymbol{e}_k.$$

Since this system is linear, the stability is determined by the eigenvalues of the matrix $\boldsymbol{M}_k = s'(y_k^*)\boldsymbol{A}_k/d_{\mathcal{C}_k} - \boldsymbol{I}_{N_k}$ restricted to the subspace $\mathbf{1}_{N_k}^{\perp}$ (since $\boldsymbol{e}_k \perp \mathbf{1}_{N_k}$). The eigenvalues of $\boldsymbol{A}_k$ are $\lambda_1 \leq \cdots \leq \lambda_{N_k}$. By the Perron-Frobenius Theorem, since the graph induced by $\boldsymbol{A}_k$ is connected, its largest eigenvalue λ_{N_k} has multiplicity one and corresponds to the eigenvector $\mathbf{1}_{N_k}$, and is thus irrelevant for $\boldsymbol{e}_k$. The eigenvalues of $\boldsymbol{M}_k$ governing the disagreement are $s'(y_k^*)\lambda_i(\boldsymbol{A}_k)/d_{\mathcal{C}_k} - 1$ for $i < N_k$. The largest real part among these is determined by λ_{N_k-1} with the maximum eigenvalue being $s'(y_k^*)\lambda_{N_k-1}(\boldsymbol{A}_k)/d_{\mathcal{C}_k} - 1 = -\alpha_k$. Thus, by the linearization principle, the system is stable if and only if $\alpha_k > 0$. □

Inter-Cell Coupling Stability When cells are interconnected, disagreements can propagate. We model this via a comparison system involving the norms of the errors. We define the *cross-coupling gain matrix* $\boldsymbol{G} \in \mathbb{R}^{M\times M}$:

$$G_{kj} = \begin{cases} -\alpha_k & \text{if } k = j \\ \dfrac{s'(y_j^*)}{d_{\mathcal{C}_k}}\sqrt{b_{kj}b_{jk}} & \text{if } k \neq j \end{cases}. \tag{5.9}$$

The diagonal entries $-\alpha_k$ represent the self-damping within each cell, while the off-diagonal entries quantify the destabilizing influence that disagreement in one cell can impose on another.

Remark 5.3.3. *Note that if a cell $\mathcal{C}_k$ has $N_k = 1$, it is trivially connected, and the internal cohesion stability condition (5.8) does not apply to that cell.*

Proposition 5.3.10 (Transversal Stability). *Let $\boldsymbol{x}^* = \boldsymbol{P}\boldsymbol{y}^*$ be a CSE resulting from a connected EP π. Assume that for all cells with $N_k \geq 2$, $\alpha_k > 0$. If the coupling gain matrix $\boldsymbol{G}$ defined in (5.9) is Hurwitz, then the full transversal dynamics is locally exponentially stable (i.e., $\boldsymbol{e}(t) \to \mathbf{0}$ exponentially).*

Proof. Since the destabilization emerging from the internal cohesion has been characterized in Proposition 5.3.9, let us focus on the inter-cell perturbation component of (5.7).

For the external influence, we decompose the perturbation $\boldsymbol{\delta}_j$ from another cell $\mathcal{C}_j$ into its average component and its disagreement component: $\boldsymbol{\delta}_j = \bar{\delta}_j \mathbf{1}_{N_j} + \boldsymbol{e}_j$. As before, $\boldsymbol{A}_{kj}$ (the $N_k \times N_j$ block of $\boldsymbol{A}$) has a constant row sum b_{kj}, thus $\boldsymbol{A}_{kj}\mathbf{1}_{N_j} = b_{kj}\mathbf{1}_{N_k}$. The projection of the influence from the average part of $\boldsymbol{\delta}_j$ is:

$$\boldsymbol{\Pi}_k^{\perp} \boldsymbol{A}_{kj}(\bar{\delta}_j \mathbf{1}_{N_j}) = \bar{\delta}_j \boldsymbol{\Pi}_k^{\perp}(\boldsymbol{A}_{kj}\mathbf{1}_{N_j}) = \bar{\delta}_j \boldsymbol{\Pi}_k^{\perp}(b_{kj}\mathbf{1}_{N_k}) = \mathbf{0}.$$

The average perturbation from other cells does not affect the internal disagreement dynamics. The only influence comes from the disagreement vectors $\boldsymbol{e}_j$:

$$\boldsymbol{\Pi}_k^{\perp} \boldsymbol{J}_{kj}\boldsymbol{\delta}_j = \frac{s'(y_j^*)}{d_{\mathcal{C}_k}} \boldsymbol{\Pi}_k^{\perp} \boldsymbol{A}_{kj}(\bar{\delta}_j \mathbf{1}_{N_j} + \boldsymbol{e}_j) = \frac{s'(y_j^*)}{d_{\mathcal{C}_k}} \boldsymbol{\Pi}_k^{\perp} \boldsymbol{A}_{kj}\boldsymbol{e}_j.$$

Combining these terms and defining the external influence as $\boldsymbol{u}_k = \sum_{j\neq k} s'(y_j^*)\boldsymbol{A}_{kj}\boldsymbol{e}_j$, we get the equation for the transversal dynamics:

$$\dot{\boldsymbol{e}}_k = \left(\frac{s'(y_k^*)}{d_{\mathcal{C}_k}} \boldsymbol{A}_k - \boldsymbol{I}_{N_k} \right) \boldsymbol{e}_k + \frac{1}{d_{\mathcal{C}_k}} \boldsymbol{\Pi}_k^{\perp} \boldsymbol{u}_k,$$

For each subsystem k, we define the degree-weighted Lyapunov function: $V_k(\boldsymbol{e}_k) = d_{\mathcal{C}_k}\|\boldsymbol{e}_k\|^2/2$. Its time derivative along the trajectories of the disagreement dynamics is:

$$\dot{V}_k(\boldsymbol{e}_k) = d_{\mathcal{C}_k}\boldsymbol{e}_k^{\top}\dot{\boldsymbol{e}}_k = \boldsymbol{e}_k^{\top}\left(s'(y_k^*)\boldsymbol{A}_k - d_{\mathcal{C}_k}\boldsymbol{I}_{N_k}\right)\boldsymbol{e}_k + \boldsymbol{e}_k^{\top}\boldsymbol{\Pi}_k^{\perp}\boldsymbol{u}_k.$$

For the first term, using the Courant-Fischer Min-Max Theorem, since $\boldsymbol{e}_k \perp \mathbf{1}_{N_k}$ it yields $\boldsymbol{e}_k^{\top}\boldsymbol{A}_k\boldsymbol{e}_k \leq \lambda_{N_k-1}(\boldsymbol{A}_k)\|\boldsymbol{e}_k\|^2$. Thus, one has

$$\begin{aligned}\boldsymbol{e}_k^{\top}(s'(y_k^*)\boldsymbol{A}_k - d_{\mathcal{C}_k}\boldsymbol{I}_{N_k})\boldsymbol{e}_k &\leq (s'(y_k^*)\lambda_{N_k-1}(\boldsymbol{A}_k) - d_{\mathcal{C}_k})\|\boldsymbol{e}_k\|^2 \\ &= -\left[d_{\mathcal{C}_k} - s'(y_k^*)\lambda_{N_k-1}(\boldsymbol{A}_k)\right]\|\boldsymbol{e}_k\|^2 = -d_{\mathcal{C}_k}\alpha_k\|\boldsymbol{e}_k\|^2.\end{aligned}$$

For the second term, Cauchy-Schwarz gives $\boldsymbol{e}_k^{\top}\boldsymbol{u}_k \leq \|\boldsymbol{e}_k\|\|\boldsymbol{u}_k\|$. We bound the input norm $\|\boldsymbol{u}_k\|$ by

$$\|\boldsymbol{u}_k\| = \left\| \sum_{j\neq k} s'(y_j^*)\boldsymbol{A}_{kj}\boldsymbol{e}_j \right\| \leq \sum_{j\neq k} s'(y_j^*)\|\boldsymbol{A}_{kj}\|_2\|\boldsymbol{e}_j\|.$$

Using the spectral norm bound for a directed adjacency matrix [127, Theorem 1], $\|\boldsymbol{A}_{kj}\|_2 \leq \sqrt{b_{kj}b_{jk}}$, we have:

$$\|\boldsymbol{u}_k\| \leq \sum_{j\neq k} s'(y_j^*)\sqrt{b_{kj}b_{jk}}\|\boldsymbol{e}_j\|.$$

This yields the differential inequality:

$$\dot{V}_k(\boldsymbol{e}_k) \leq -d_{\mathcal{C}_k}\alpha_k\|\boldsymbol{e}_k\|^2 + \|\boldsymbol{e}_k\| \sum_{j\neq k} s'(y_j^*)\sqrt{b_{kj}b_{jk}}\|\boldsymbol{e}_j\|. \tag{5.10}$$

Let $z_k(t) = \|\boldsymbol{e}_k(t)\|$. Substituting z_k in (5.10) (and dividing by z_k for $z_k \neq 0$) gives a system of linear differential inequalities for the vector of norms $\boldsymbol{z}$:

$$\dot{z}_k \leq -\alpha_k z_k + \sum_{j\neq k} \frac{s'(y_j^*)}{d_{\mathcal{C}_k}}\sqrt{b_{kj}b_{jk}}z_j,$$

[127] "On the Spectral Radius of a Directed Graph". Jaroslaw Kwapisz. (1996).

which can be written as $\dot{\boldsymbol{z}} \leq \boldsymbol{G}\boldsymbol{z}$. By the Comparison Lemma, if the linear system $\dot{\boldsymbol{\eta}} = \boldsymbol{G}\boldsymbol{\eta}$ is exponentially stable (i.e., if $\boldsymbol{G}$ is Hurwitz), then $\boldsymbol{z}(t)$ converges exponentially to zero. This establishes the local exponential stability of the transversal dynamics to the synchronization manifold $\mathcal{S}_\pi$. □

Main Stability Theorem for CSEs

We combine the analyses of the quotient and transversal dynamics to provide sufficient conditions for the full local stability of the equilibrium.

Theorem 5.3.11. *Let $\boldsymbol{x}^* = \boldsymbol{P}\boldsymbol{y}^*$ be a CSE resulting from a connected EP π. Assume the signal function s is continuously differentiable in a neighborhood of each component of $\boldsymbol{y}^*$. If the following two conditions hold:*

*i) **Quotient Stability:** The quotient Jacobian matrix $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$ defined in (5.6) is Hurwitz.*

*ii) **Transversal Stability:** For every cell $\mathcal{C}_k$ with cardinality $N_k \geq 2$, the internal decay rate α_k defined in (5.8) is positive, and the cross-coupling gain matrix $\boldsymbol{G}$ defined in (5.9) is Hurwitz.*

Then, the equilibrium $\boldsymbol{x}^*$ is locally exponentially stable.

Conversely, if $\boldsymbol{x}^*$ is locally exponentially stable, then $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$ is Hurwitz and $\alpha_k > 0$ for all k.

Proof. The eigenvalues of the full Jacobian $\boldsymbol{J}(\boldsymbol{x}^*)$ determine local stability. By Lemma 5.3.7, the set of eigenvalues of $\boldsymbol{J}_{\text{quot}}(\boldsymbol{y}^*)$ is a subset of the eigenvalues of $\boldsymbol{J}(\boldsymbol{x}^*)$. Condition (i) ensures that these M eigenvalues have strictly negative real parts.

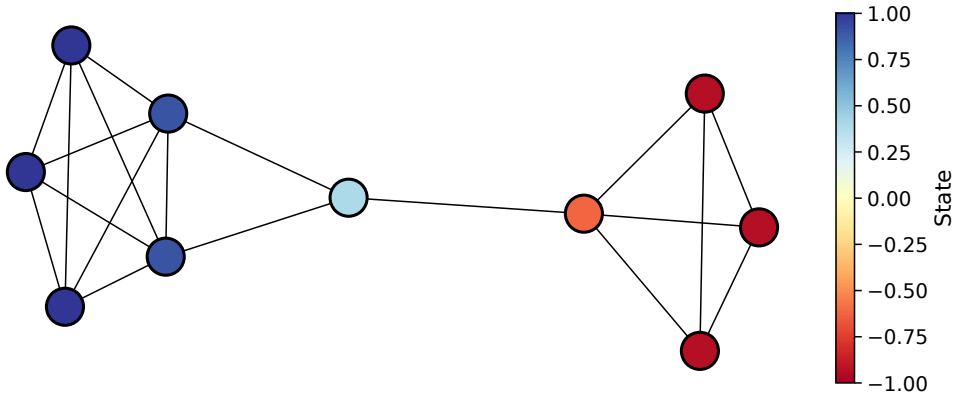
The remaining $N - M$ eigenvalues govern the dynamics of the disagreement subspace (transversal dynamics). Condition (ii) ensures, via Proposition 5.3.10, that the transversal dynamics is exponentially stable. Since both invariant subspaces are exponentially stable, the full system is locally exponentially stable.

Conversely, if $\boldsymbol{J}(\boldsymbol{x}^*)$ is Hurwitz, then its subset of eigenvalues corresponding to $\boldsymbol{J}_{\text{quot}}$ must be stable (Proposition 5.3.8), and the diagonal blocks corresponding to internal modes must be stable (Proposition 5.3.9), implying $\alpha_k > 0$ for every cell with $N_k \geq 2$. □

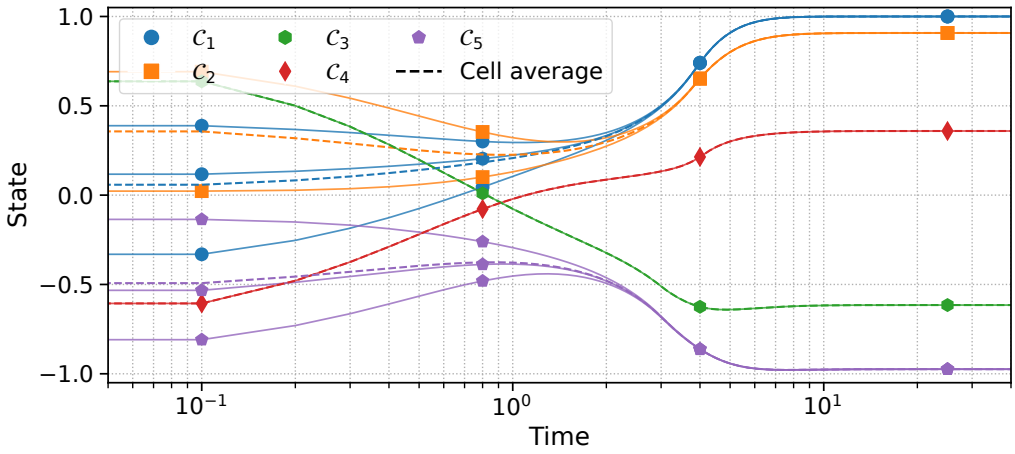
Remark 5.3.4 (Necessity of the Transversal Stability Condition)**.** *While condition (i) is necessary and sufficient for quotient stability, condition (ii) is generally only a* sufficient *condition for the stability of the full transversal dynamics. This conservatism arises in the derivation of* (5.3.3)*, where we upper-bound the influence of external disagreements by taking the absolute values of the interaction terms. This bounding process inherently neglects the potential cancellations of errors arriving from different neighbors.*

In network topologies with strong internal coupling, these cancellations are deeply related to the concept of frustration *in statistical mechanics (e.g., the Ising model). The graph topology itself might structurally suppress the propagation of disagreements, enforcing stability even when the comparison matrix $\boldsymbol{G}$ is not Hurwitz. In theory, one could formulate an exact necessary and sufficient condition by precisely tracking these sign cancellations via a modified matrix. However, evaluating such a condition is computationally intractable. Finding the exact stability threshold under structural frustration essentially maps to the Ising frustration problem (i.e., finding the ground state of a frustrated spin glass), which is well-known to be NP-hard (see [128]). Thus, relying on the Hurwitz property of the upper-bound matrix $\boldsymbol{G}$ provides a practical and computationally efficient, but more conservative, stability certificate.*

[128] "On the Computational Complexity of Ising Spin Glass Models". F. Barahona. (1982).



(a) Asymptotic state of the system.



(b) Trajectory of the system with equitable partition in log scale. Each node in the cell has the same color as the cell. The average of each cell is plotted in a dashed line.

Figure 5.4: Numerical validation of convergence to a stable CSE on the graph from Fig. 5.2a. The simulation uses $s(x) = \min(1, \max(-1, Kx))$ with $K = 1.2$. **(Top)** The final equilibrium state, where agents within each of the five cells converge to a distinct common value, forming a stable CSE. **(Bottom)** Agent trajectories (solid lines) reveal a clear two-stage dynamic. First, they rapidly contract onto their respective cell averages (dashed lines), demonstrating the system's *transversal stability*. Subsequently, these cell averages converge more slowly to their final equilibrium values, governed by the stable quotient dynamics, thereby confirming *quotient stability*. This provides a concrete validation of the stability decomposition established in Theorem 5.3.11.

Remark 5.3.5 (Connected EP assumption)**.** *The assumption that the EP π is connected ensures that each induced subgraph $\mathcal{G}_k$ is connected. If the connectivity assumption on the partition is dropped, a cell $\mathcal{C}_k$ with $N_k \geq 2$ might have $b_{kk} = 0$ (no internal edges). In this scenario, $\boldsymbol{D}_k = \mathbf{0}$ and $\boldsymbol{L}_k = -\boldsymbol{A}_k$. The stability matrix for transversal modes becomes $\boldsymbol{M}_k = s'(y_k^*)\boldsymbol{A}_k - \boldsymbol{I}_{N_k}$. Stability then requires $s'(y_k^*)\lambda_j(\boldsymbol{A}_k) < 1$ for all eigenvalues $\lambda_j(\boldsymbol{A}_k)$ of $\boldsymbol{A}_k$ corresponding to transversal modes.*

Remark 5.3.6 (Relaxation of the smoothness of s)**.** *The previous propositions and theorem assume that the signal function s is continuously differentiable near the equilibrium $\boldsymbol{x}^*$. While this facilitates a tractable proof using standard linearization, the result can theoretically be extended to non-smooth, Lipschitz continuous signals (e.g., piecewise linear functions) using non-smooth analysis tools such as the Clarke generalized Jacobian [129, Section 1.3]. In that framework, the scalar derivative $s'(y_k^*)$ in conditions (5.6) and (5.8) would be replaced by elements of the generalized gradient set $\partial s(y_k^*)$. Stability would then require that the quotient and transversal conditions hold for all values within this set.*

A numerical simulation validates the stability decomposition established in Theorem 5.3.11. Figure 5.4 shows the formation of a stable CSE on the graph previously introduced in Figure 5.2a. The signal function and initial conditions were chosen to satisfy the stability conditions and permit a clustered equilibrium. As detailed in the caption, the trajectories clearly exhibit the two distinct timescales predicted by the theory, providing a concrete illustration of both transversal and quotient stability.

5.4 Approximated Local Synchronization: an Input-to-State Stability Concept

The preceding analysis centered on *perfect local synchronization*, a phenomenon constrained by the existence of an equitable partition. While this provided a structural condition for the formation of exact clusters, such perfect symmetries are rare in many real-world networks. This raises a critical and more general question: can cohesive clusters persist in networks that lack these rigid structural properties, and if so, what determines their internal coherence?

This section addresses this question by formalizing the concept of *approximated local synchronization*. In this regime, agents within a cluster do not converge to a single point, but the disagreement among them remains ultimately bounded. This represents a shift from analyzing an equilibrium property (perfect agreement) to a robustness property (bounded disagreement). Instead of asking whether internal disagreement vanishes, we ask how resilient a cluster's internal cohesion is to disruptive influences.

To rigorously analyze this phenomenon, we adopt a perspective from robust control theory. We formally treat a candidate cluster as a nominal dynamical system that is perpetually forced by an external input signal originating from the remainder of the network. The framework of *Input-to-State Stability (ISS)* is the natural tool for this analysis. It provides a formal means to quantify the relationship between the magnitude of the external perturbations (the "input") and the ultimate bound on the internal cluster disagreement (the "state").

Our first step is therefore to isolate the dynamics of a given cluster and precisely decompose it into its ideal, internal dynamics and the perturbation terms that prevent perfect synchronization. This decomposition will allow us to define the state and input variables for the subsequent ISS analysis.

[129] *Nonsmooth Vector Functions and Continuous Optimization*. Vaithilingam Jeyakumar. (2008).

5.4.1 ISS Analysis of Internal Cluster Disagreement

We now apply this framework to analyze the dynamics within a connected subgraph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ that contains N' nodes. Our goal is to express the dynamics of the cluster $\mathcal{V}'$ in the form of an ISS system, where the "state" is the internal disagreement within the cluster and the "input" is the influence from the rest of the network, $\mathcal{V} \setminus \mathcal{V}'$.

Let $\boldsymbol{S} \in \mathbb{R}^{N' \times N}$ be the selection matrix (Definition 2.2.23) that maps the full state vector $\boldsymbol{x}$ to the subgraph state vector $\boldsymbol{x}' := \boldsymbol{S}\boldsymbol{x}$. Specifically, for $j \in \mathcal{V}'$, $\boldsymbol{S}_{ij} = 1$ for its corresponding index $i \in \mathcal{V}$ and 0 otherwise. Note that $\boldsymbol{S}\boldsymbol{S}^\top = \boldsymbol{I}_{N'}$ and $\boldsymbol{S}^\top\boldsymbol{S} = \mathrm{diag}(\mathbb{I}_{\mathcal{V}'})$, where $\mathbb{I}_{\mathcal{V}'} \in \mathbb{R}^N$ is the indicator vector of $\mathcal{V}' \subseteq \mathcal{V}$. i.e., $[\mathbb{I}_{\mathcal{V}'}]_i = 1$ if node $i \in \mathcal{V}$ is in $\mathcal{V}'$ and 0 otherwise.

We can decompose the term $\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x})$ based on the subgraph structure. Let $\boldsymbol{A}_{\text{in}} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{S}^\top \in \mathbb{R}^{N \times N}$ be the adjacency matrix of the induced subgraph $\mathcal{G}'$. Let $\boldsymbol{A}_{\text{ext}}$ be the matrix representing connections between $\mathcal{V}'$ and $\mathcal{V} \setminus \mathcal{V}'$. Then, for the nodes in $\mathcal{V}'$, the input from the network can be split into an internal and an external part:

$$\boldsymbol{S}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{A}_{\text{in}}\boldsymbol{s}(\boldsymbol{x}') + \boldsymbol{S}\boldsymbol{A}_{\text{ext}}\boldsymbol{s}(\boldsymbol{x}). \tag{5.11}$$

Let us define the following:

- $\boldsymbol{D}_{\text{in}} = \mathrm{diag}(d_i^{\text{in}})_{i \in \mathcal{V}'}$, where $d_i^{\text{in}} = \sum_{j \in \mathcal{V}'} a_{ij}$, is the diagonal matrix of internal degrees for nodes in $\mathcal{G}'$.
- $\widetilde{\boldsymbol{D}}_{\text{in}} = \boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^\top \in \mathbb{R}^{N' \times N'}$ is the diagonal matrix containing the total degrees d_i for nodes $i \in \mathcal{V}'$ in the original graph $\mathcal{G}$. Note that $\widetilde{\boldsymbol{D}}_{\text{in}}^{-1} = \boldsymbol{S}\boldsymbol{D}^{-1}\boldsymbol{S}^\top$. Since $\boldsymbol{S}^\top\boldsymbol{S}$ is diagonal and $\boldsymbol{S}\boldsymbol{S}^\top = \boldsymbol{I}_{N'}$, one has $\widetilde{\boldsymbol{D}}_{\text{in}}^{-1}\boldsymbol{S} = \boldsymbol{S}\boldsymbol{D}^{-1}\boldsymbol{S}^\top\boldsymbol{S} = \boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{S}\boldsymbol{D}^{-1} = \boldsymbol{S}\boldsymbol{D}^{-1}$.

Substituting these definitions into (5.11) and rearranging allows us to express the dynamics as an ideal internal system plus two distinct perturbation terms:

$$\begin{aligned}\dot{\boldsymbol{x}}' &= \boldsymbol{D}_{\text{in}}^{-1}\boldsymbol{A}_{\text{in}}\boldsymbol{s}(\boldsymbol{x}') - \boldsymbol{x}' + \left(\widetilde{\boldsymbol{D}}_{\text{in}}^{-1} - \boldsymbol{D}_{\text{in}}^{-1}\right)\boldsymbol{A}_{\text{in}}\boldsymbol{s}(\boldsymbol{x}') + \widetilde{\boldsymbol{D}}_{\text{in}}^{-1}\boldsymbol{S}\boldsymbol{A}_{\text{ext}}\boldsymbol{s}(\boldsymbol{x}) \\ &:= \underbrace{\left(\boldsymbol{D}_{\text{in}}^{-1}\boldsymbol{A}_{\text{in}}\boldsymbol{s}(\boldsymbol{x}') - \boldsymbol{x}'\right)}_{\text{Ideal Internal Dynamics}} + \underbrace{\boldsymbol{p}(\boldsymbol{x})}_{\text{Total Perturbation}}.\end{aligned} \tag{5.12}$$

The total perturbation $\boldsymbol{p}(\boldsymbol{x})$ originates from two sources:

- *External Perturbation*: Influence from neighbors outside the cluster, given by $\widetilde{\boldsymbol{D}}_{\text{in}}^{-1}\boldsymbol{S}\boldsymbol{A}_{\text{ext}}\boldsymbol{s}(\boldsymbol{x})$.
- *Internal Perturbation*: A more subtle effect arising from the mismatch between a node's total degree and its internal degree, given by $\left(\widetilde{\boldsymbol{D}}_{\text{in}}^{-1} - \boldsymbol{D}_{\text{in}}^{-1}\right)\boldsymbol{A}_{\text{in}}\boldsymbol{s}(\boldsymbol{x}')$.

Remark 5.4.1. *Note that this internal perturbation was missing from the EP analysis (Proof of Theorem 5.3.11). This is because this difference is non-zero only if $d_i \neq d_i^{\text{in}}$, which by definition means that agent i has at least one connection outside the cluster $\mathcal{V}'$. In other words, this perturbation only affects the boundary nodes of the cluster. In the case of $\mathcal{V}'$ being a cell of an EP, every node is on the boundary of the cluster, thus the deviation from the internal degree is the same for all nodes. Indeed, $(\widetilde{\boldsymbol{D}}_{\text{in}}^{-1} - \boldsymbol{D}_{\text{in}}^{-1}) = (1/d_i - 1/d_i^{\text{in}})\mathbf{I}_{N'}$ for any $i \in \mathcal{V}'$.*

A key insight of the internal consensus dynamics is that the system's internal disagreement is *insensitive to any uniform perturbation that pushes all nodes in the cluster equally*. The component of the perturbation that drives disagreement is its non-uniform part. We therefore define the *residual perturbation* as the component of $\boldsymbol{p}(\boldsymbol{x})$ that is orthogonal to the fully synchronized manifold of the cluster:

$$\tilde{\boldsymbol{p}}(\boldsymbol{x}) := \boldsymbol{p}(\boldsymbol{x}) - \bar{p}(\boldsymbol{x})\mathbf{1}_{N'},$$

where $\bar{p}(\boldsymbol{x}) = (\sum_{i\in\mathcal{V}'} d_i^{\text{in}}[p(\boldsymbol{x})]_i)/(\sum_{i\in\mathcal{V}'} d_i^{\text{in}})$ is the weighted average of the perturbation over the cluster.

The norm of the residual perturbation,

$$\|\tilde{\boldsymbol{p}}(\boldsymbol{x})\|^2_{\boldsymbol{D}_{\text{in}}} = \sum_{i=1}^{N'} d_i^{\text{in}}([p(\boldsymbol{x})]_i - \bar{p}(\boldsymbol{x}))^2,$$

is a weighted variance of the total perturbation's effect on the subgraph. The closest $\|\tilde{\boldsymbol{p}}(\boldsymbol{x})\|_{\boldsymbol{D}_{\text{in}}}$ is to 0, the most uniform is the perturbation across the subgraph $\mathcal{G}'$. We can define an upper bound for this norm as:

$$\|\tilde{\boldsymbol{p}}(\boldsymbol{x})\|_{\boldsymbol{D}_{\text{in}}} \leq \sup_{\boldsymbol{x}\in\mathcal{X}} \|\tilde{\boldsymbol{p}}(\boldsymbol{x})\|_{\boldsymbol{D}_{\text{in}}} := P^{\text{sup}}.$$

We can now state the main result of this section.

Theorem 5.4.1 (ISS of Synchronized Clusters)**.** *Consider a connected induced subgraph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ of the graph $\mathcal{G}$. Let us define $\alpha_{\text{in}} := 1 - K \max_{i\in\{1,\dots,N-1\}} |\lambda_i(\boldsymbol{D}_{\text{in}}^{-1}\boldsymbol{A}_{\text{in}})|$.*

If the internal stability condition $\alpha_{\text{in}} > 0$ holds, then the disagreement within the subgraph $\mathcal{G}'$ is input-to-state stable with respect to the residual perturbation $\tilde{\boldsymbol{p}}(\boldsymbol{x})$.

Specifically, for the disagreement vector $\boldsymbol{e} = \boldsymbol{x}' - \bar{x}'\mathbf{1}_{N'}$, where $\bar{x}' = (\sum_{i\in\mathcal{V}'} d_i^{\text{in}} x_i)/(\sum_{i\in\mathcal{V}'} d_i^{\text{in}})$, there exist a $\mathcal{KL}$ function β and a $\mathcal{K}$ function γ, given by

$$\beta(r,t) = re^{-\alpha_{\text{in}}t} \quad \textit{and} \quad \gamma(r) = \frac{r}{\alpha_{\text{in}}},$$

such that the following ISS inequality holds for all $t \geq 0$:

$$\|\boldsymbol{e}(t)\|_{\boldsymbol{D}_{\text{in}}} \leq \beta(\|\boldsymbol{e}(0)\|_{\boldsymbol{D}_{\text{in}}}, t) + \gamma\left(\sup_{0\leq\tau\leq t} \|\tilde{\boldsymbol{p}}(\boldsymbol{x}(\tau))\|_{\boldsymbol{D}_{\text{in}}}\right).$$

Moreover, the norm of the disagreement $\|\boldsymbol{e}\|_{\boldsymbol{D}_{\text{in}}}$ is ultimately bounded by $P^{\text{sup}}/\alpha_{\text{in}}$.

Proof. The proof starts by considering the ISS-Lyapunov function candidate:

$$V(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}' - \bar{x}'\mathbf{1}_{N'})^\top \boldsymbol{D}_{\text{in}}(\boldsymbol{x}' - \bar{x}'\mathbf{1}_{N'}) = \frac{1}{2}\boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\boldsymbol{e} = \frac{1}{2}\|\boldsymbol{e}\|^2_{\boldsymbol{D}_{\text{in}}}.$$

From the dynamics in (5.12) and following similar steps to the proof of Theorem 4.6.1, the time derivative of $V(\boldsymbol{x})$ along the trajectories of the system is:

$$\dot{V}(\boldsymbol{x}) = \boldsymbol{e}^\top \boldsymbol{A}_{\text{in}}\boldsymbol{s}(\boldsymbol{x}') - \boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\boldsymbol{x}' + \boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\boldsymbol{p}(\boldsymbol{x}).$$

The internal dynamics term is bounded as (from the proof of Theorem 4.6.1):

$$\boldsymbol{e}^\top \boldsymbol{A}_{\text{in}}\boldsymbol{s}(\boldsymbol{x}') - \boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\boldsymbol{x}' \leq -2\alpha_{\text{in}}V(\boldsymbol{x}).$$

The perturbation term $\boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\boldsymbol{p}(\boldsymbol{x})$ is analyzed by decomposing the perturbation vector $\boldsymbol{p}(\boldsymbol{x})$ into its $\boldsymbol{D}_{\text{in}}$-weighted average, $\bar{p}(\boldsymbol{x})\mathbf{1}_{N'}$, and the residual part, $\tilde{\boldsymbol{p}}(\boldsymbol{x})$. Thus, $\boldsymbol{p}(\boldsymbol{x}) = \bar{p}(\boldsymbol{x})\mathbf{1}_{N'} + \tilde{\boldsymbol{p}}(\boldsymbol{x})$.

Substituting this decomposition into the perturbation term yields:

$$\boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\boldsymbol{p}(\boldsymbol{x}) = \boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\left(\bar{p}(\boldsymbol{x})\mathbf{1}_{N'} + \tilde{\boldsymbol{p}}(\boldsymbol{x})\right) = \bar{p}(\boldsymbol{x})\boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\mathbf{1}_{N'} + \boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\tilde{\boldsymbol{p}}(\boldsymbol{x}) = \boldsymbol{e}^\top \boldsymbol{D}_{\text{in}}\tilde{\boldsymbol{p}}(\boldsymbol{x}).$$

Since $\boldsymbol{e}^\top \boldsymbol{D}_{\text{in}} \mathbf{1}_{N'} = 0$ by definition of the weighted average $\bar{x}'$, from Lemma 4.2.4, the first term vanishes.

Using the Cauchy-Schwarz inequality for the $\boldsymbol{D}_{\text{in}}$-weighted inner product $\langle \mathbf{u}, \mathbf{v} \rangle_{\boldsymbol{D}_{\text{in}}} = \mathbf{u}^\top \boldsymbol{D}_{\text{in}} \mathbf{v}$:

$$\boldsymbol{e}^\top \boldsymbol{D}_{\text{in}} \tilde{\boldsymbol{p}}(\boldsymbol{x}) = \langle \boldsymbol{e}, \tilde{\boldsymbol{p}}(\boldsymbol{x}) \rangle_{\boldsymbol{D}_{\text{in}}} \leq \|\boldsymbol{e}\|_{\boldsymbol{D}_{\text{in}}} \|\tilde{\boldsymbol{p}}(\boldsymbol{x})\|_{\boldsymbol{D}_{\text{in}}}.$$

Combining these, the differential inequality for $V(\boldsymbol{x})$ becomes:

$$\dot{V}(\boldsymbol{x}) \leq -2\alpha_{\text{in}} V(\boldsymbol{x}) + \|\boldsymbol{e}\|_{\boldsymbol{D}_{\text{in}}} \|\tilde{\boldsymbol{p}}(\boldsymbol{x})\|_{\boldsymbol{D}_{\text{in}}}.$$

Let $y = \|\boldsymbol{e}\|_{\boldsymbol{D}_{\text{in}}}$. Then $V(\boldsymbol{x}) = \frac{1}{2} y^2$ and its time derivative is $\dot{V}(\boldsymbol{x}) = y\dot{y}$. Substituting these into the inequality gives:

$$y\dot{y} \leq -\alpha_{\text{in}} y^2 + y \|\tilde{\boldsymbol{p}}(\boldsymbol{x})\|_{\boldsymbol{D}_{\text{in}}}.$$

For any $y > 0$, i.e., $\|\boldsymbol{e}\|_{\boldsymbol{D}_{\text{in}}} > 0$, we can divide by y:

$$\dot{y} \leq -\alpha_{\text{in}} y + \|\tilde{\boldsymbol{p}}(\boldsymbol{x})\|_{\boldsymbol{D}_{\text{in}}} = -\alpha_{\text{in}} y + u,$$

where $u = \|\tilde{\boldsymbol{p}}(\boldsymbol{x})\|_{\boldsymbol{D}_{\text{in}}}$. This is a linear differential inequality $\dot{y} \leq -\alpha_{\text{in}} y + u$, which also holds trivially when $y = 0$.

Applying Gronwall's inequality, the solution is bounded by:

$$\begin{aligned} y(t) &\leq y(0) e^{-\alpha_{\text{in}} t} + \int_0^t e^{-\alpha_{\text{in}}(t-\tau)} u(\tau) d\tau \leq y(0) e^{-\alpha_{\text{in}} t} + \left(\sup_{0 \leq \tau \leq t} u(\tau) \right) \int_0^t e^{-\alpha_{\text{in}}(t-\tau)} d\tau \\ &= y(0) e^{-\alpha_{\text{in}} t} + \frac{1}{\alpha_{\text{in}}} \left(\sup_{0 \leq \tau \leq t} u(\tau) \right) (1 - e^{-\alpha_{\text{in}} t}). \end{aligned}$$

Since $(1 - e^{-\alpha_{\text{in}} t}) \leq 1$ for $t \geq 0$, we have the simpler bound:

$$y(t) \leq y(0) e^{-\alpha_{\text{in}} t} + \frac{1}{\alpha_{\text{in}}} \sup_{0 \leq \tau \leq t} u(\tau).$$

Substituting back $y(t) = \|\boldsymbol{e}(t)\|_{\boldsymbol{D}_{\text{in}}}$ yields the ISS inequality:

$$\|\boldsymbol{e}(t)\|_{\boldsymbol{D}_{\text{in}}} \leq \|\boldsymbol{e}(0)\|_{\boldsymbol{D}_{\text{in}}} e^{-\alpha_{\text{in}} t} + \frac{1}{\alpha_{\text{in}}} \sup_{0 \leq \tau \leq t} \|\tilde{\boldsymbol{p}}(\boldsymbol{x}(\tau))\|_{\boldsymbol{D}_{\text{in}}}.$$

with $\beta(r, t) = r e^{-\alpha_{\text{in}} t}$ and $\gamma(r) = r / \alpha_{\text{in}}$.

For the ultimate bound, taking the limit sup as $t \to \infty$:

$$\limsup_{t \to \infty} \|\boldsymbol{e}(t)\|_{\boldsymbol{D}_{\text{in}}} \leq \limsup_{t \to \infty} \left(\|\boldsymbol{e}(0)\|_{\boldsymbol{D}_{\text{in}}} e^{-\alpha_{\text{in}} t} + \frac{1}{\alpha_{\text{in}}} \sup_{0 \leq \tau \leq t} \|\tilde{\boldsymbol{p}}(\boldsymbol{x}(\tau))\|_{\boldsymbol{D}_{\text{in}}} \right).$$

Since $\alpha_{\text{in}} > 0$, the first term goes to zero, leaving:

$$\limsup_{t \to \infty} \|\boldsymbol{e}(t)\|_{\boldsymbol{D}_{\text{in}}} \leq \frac{1}{\alpha_{\text{in}}} \sup_{t \geq 0} \|\tilde{\boldsymbol{p}}(\boldsymbol{x}(t))\|_{\boldsymbol{D}_{\text{in}}} \leq \frac{P^{\text{sup}}}{\alpha_{\text{in}}}.$$

This concludes the proof. □

Theorem 5.4.1 states that a sufficiently cohesive cluster can maintain its internal agreement, even when perpetually influenced by the rest of the network. The theorem formalizes this by demonstrating that the cluster's internal disagreement is ISS. This implies that while initial disagreement within the cluster decays exponentially, persistent external influence results in a finite ultimate bound on this disagreement. Thus, Theorem 5.4.1 provides a tool for analyzing the stability of NFSE, interpreting the network as an interconnection of robust and internally stable clusters.

An important consequence of this framework is the identification of *variance* (or non-uniformity) of the external perturbation being the true cause of internal disagreement. A uniform influence, regardless of its magnitude, would not prevent the cluster from achieving perfect internal consensus asymptotically. It is therefore the *heterogeneity* of external signals, which causes boundary nodes to be pulled in different directions, that limits local exact synchronization. This explains the clustering obtained in [68] (in Appendix A.2 and in Appendix A.3), where a uniform environmental signal becomes an effectively non-uniform perturbation due to heterogeneous agent gains, sustaining disagreement as our framework predicts.

Remark 5.4.2 (A Unifying Principle of Stability)**.** *The condition for a cluster's internal stability, $\alpha_{\text{in}} > 0$, is structurally identical to the condition for global exponential stability in Theorem 4.6.1. This reveals a scale-free principle: a group of agents, whether the entire network or a single cluster, maintains cohesion if its internal connectivity is strong enough to overcome the maximal destabilizing effect of the nonlinearity, a principle elegantly quantified by the eigenvalues of the relevant subgraph dynamics.*

Remark 5.4.3 (A Practical Bound on Perturbation)**.** *The ultimate bound on disagreement depends on the maximum residual perturbation, P^{sup}. This term can be bounded by the structural properties of the cluster's boundary:*

$$P^{\text{sup}} \leq 2 \sum_{i \in \mathcal{V}'} d_i^{\text{in}} \left(\frac{d_i^{\text{ext}}}{d_i} \right)^2,$$

where $d_i^{\text{ext}} = d_i - d_i^{\text{in}}$ is the number of external connections for node $i \in \mathcal{V}'$. This provides a concrete, computable upper bound on the asymptotic synchronization error within a cluster, based entirely on the network topology.

Remark 5.4.4 (Unifying the EP and ISS Frameworks)**.** *The ISS framework for approximated local synchronization provides a powerful generalization that contains the perfect local synchronization analysis as a special, ideal case. To see this, consider a cluster $\mathcal{V}'$ that is a cell $\mathcal{C}_k$ of an equitable partition π. Let us analyze the perturbation term $\boldsymbol{p}(\boldsymbol{x})$ from (5.12) for a state $\boldsymbol{x}$ on the Cell Synchronized Manifold S_π, where $x_j = y_l$ for all $j \in \mathcal{C}_l$.*

For any agent $i \in \mathcal{C}_k$, the EP properties imply that its total degree is a constant $d_i = d_{\mathcal{C}_k}$ and its internal degree is a constant $d_i^{\text{in}} = b_{kk}$. The perturbation acting on agent i consists of two components:

(i) ***Internal Perturbation:*** *The term $(\frac{1}{d_i} - \frac{1}{d_i^{\text{in}}}) \sum_{j \in \mathcal{C}_k} a_{ij} s(x_j)$ simplifies to $(\frac{1}{d_{\mathcal{C}_k}} - \frac{1}{b_{kk}}) b_{kk} s(y_k)$. This value is identical for every agent $i \in \mathcal{C}_k$.*

(ii) ***External Perturbation:*** *The term $\frac{1}{d_i} \sum_{j \notin \mathcal{C}_k} a_{ij} s(x_j)$ can be grouped by cells, becoming*

$$\frac{1}{d_{\mathcal{C}_k}} \sum_{l \neq k} \sum_{j \in \mathcal{C}_l} a_{ij} s(y_l).$$

By the EP definition, this simplifies to $\frac{1}{d_{\mathcal{C}_k}} \sum_{l \neq k} b_{kl} s(y_l)$, which is also identical for every agent $i \in \mathcal{C}_k$.

[68] "Analysis of an Opinion Dynamics Model Coupled with an External Environmental Dynamics". Anthony Couthures et al. (2024).

Since both components of the perturbation are uniform across all agents in the cell, the total perturbation vector $\boldsymbol{p}(\boldsymbol{x})$ is constant for all agents within C_k. Consequently, the residual perturbation $\tilde{\boldsymbol{p}}(\boldsymbol{x})$*, which is the component of $\boldsymbol{p}(\boldsymbol{x})$ orthogonal to the consensus subspace of the cell, is identically zero.*

According to 5.4.1, the internal cell disagreement is ISS with respect to $\tilde{\boldsymbol{p}}(\boldsymbol{x})$. With a zero input, the ISS inequality guarantees that the norm of the internal disagreement converges asymptotically to zero. This is precisely the definition of perfect local synchronization*. Thus, the ISS analysis correctly predicts that a cell forming part of an equitable partition will maintain perfect internal cohesion, elegantly unifying the two primary concepts of this chapter.*

Remark 5.4.5 (State-Dependency and Absolute Bounding)**.** *A reader familiar with classical ISS theory might note that the residual perturbation $\tilde{\boldsymbol{p}}(\boldsymbol{x})$ is not strictly exogenous, as it inherently depends on the internal cluster state $\boldsymbol{x}'$. In unbounded systems, this state-dependency could lead to circular amplification. However, because our system's trajectories are strictly confined to the forward-invariant compact hypercube $\mathcal{X} = [-1, 1]^N$ (Proposition 3.3.1), $\tilde{\boldsymbol{p}}(\boldsymbol{x}(t))$ is structurally guaranteed to be bounded by the absolute supremum $P^{\sup}$. Therefore, we can treat $\tilde{\boldsymbol{p}}(\boldsymbol{x}(t))$ as a bounded, unknown disturbance. The ISS inequality thus serves its intended purpose: mapping the worst-case topological heterogeneity of the graph ($P^{\sup}$) directly to the maximum persistent disagreement at equilibrium.*

5.4.2 Numerical Illustration: Robust Clustering in the Karate Club Network

To demonstrate our theoretical framework in action, we apply it to Zachary's Karate Club network [130]. As detailed in the caption of Figure 5.5, the simulation parameters are chosen specifically to violate the conditions for global synchronization, creating an environment where clustering is the expected outcome.

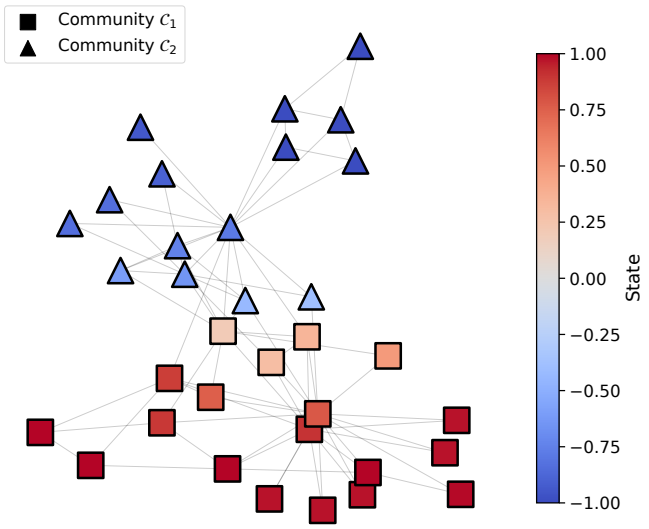
As predicted by our analysis, the network forgoes global consensus and instead converges to a stable NFSE. The final state of the network, visualized in Figure 5.5a, shows a clear partitioning of agents' opinions. The states align perfectly with the two ground-truth communities, forming two internally synchronized but mutually opposed clusters. This outcome confirms that, when the sharp threshold is crossed, the system can transition into a state of structured, local agreement.

This clustered equilibrium provides a perfect setting for validating the core predictions of our ISS analysis. In Figure 5.5b, we treat each community as a subsystem and plot the time evolution of its internal disagreement (the ISS state) and the residual perturbation it receives from the opposing cluster (the ISS input). The resulting dynamics are a standard illustration of Input-to-State Stability:

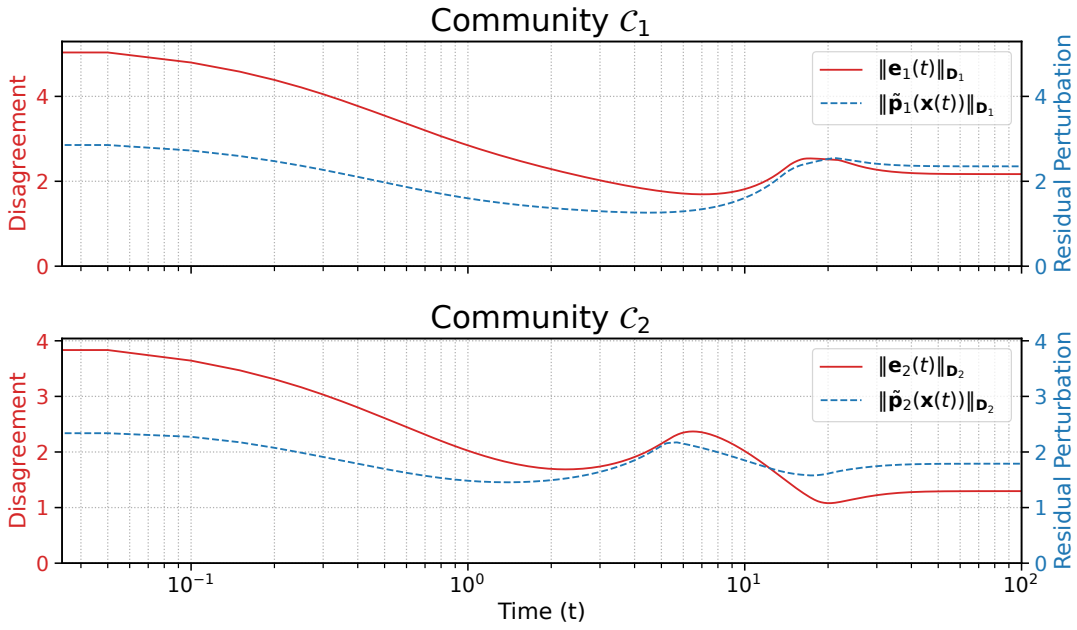
- The internal disagreement (red curve) exponentially decays from its initial value, demonstrating the cluster's internal stability ($\alpha_{\text{in}} > 0$).
- However, the disagreement does not converge to zero. It is sustained by the persistent, non-uniform perturbation from the other cluster (blue curve), settling to a non-zero ultimate bound.

This behavior, of exponential decay to a non-zero bound sustained by a persistent input, is precisely the outcome guaranteed by the ISS inequality in Theorem 5.4.1. The simulation thus provides a concrete validation of our theoretical results.

[130] "An Information Flow Model for Conflict and Fission in Small Groups". Wayne W. Zachary. (1977).



(a) Final equilibrium state of the network. Node shapes (triangle/square) indicate the two strong communities: $\mathcal{V}_1$ and $\mathcal{V}_2$. The communities were obtained using [88]. Node colors represent the final state value x_i^*, demonstrating local synchronization within clusters but global disagreement between them.



(b) ISS analysis for each community. For $k \in \{1, 2\}$, the internal disagreement $\|\boldsymbol{e}_k(t)\|_{\boldsymbol{D}_k}$ (the ISS "state", in red) decays exponentially from its initial condition and converges to a non-zero ultimate bound. This bound is sustained by the persistent residual perturbation $\|\tilde{\boldsymbol{p}}_k(\boldsymbol{x}(t))\|_{\boldsymbol{D}_k}$ from the opposing cluster (the ISS "input," in blue), precisely as predicted by Theorem 5.4.1.

Figure 5.5: Numerical validation of the ISS framework on Zachary's Karate Club network, a canonical example of a social network with a strong, empirically verified community structure. The simulation uses a signal function $s(x) = \max(-1, \min(1, Kx))$ with a gain $K = 1.2$ chosen to violate the sharp threshold for global synchronization ($K\lambda_{N-1} > 1$), forcing the emergence of a stable, clustered equilibrium (NFSE).

5.5 Cluster Deviation Stability

Theorem 5.4.1 established the input-to-state stability of intra-cell disagreement, focusing on the deviation from the cell's own weighted average state. This provides a measure of internal coherence. We now shift our attention to a different, yet related, aspect of cell stability: the deviation of individual agents within a cell from a common global reference point, c. This reference point is assumed to be a fixed point of the agents' signal function s.

Proposition 5.5.1 below explores the stability of this deviation, specifically using the maximum absolute deviation $\delta(t)$ within a subset of agents $\mathcal{V}'$ as the state. The input in this context is defined as the maximum influence exerted on any agent in the cell by its neighbors outside the cell, normalized by its total degree. This proposition provides conditions under which this cell deviation is also input-to-state stable, offering guarantees on how closely agents within a cell can track a desired global state c in the presence of external influences from other parts of the network.

Proposition 5.5.1. *Let $\mathcal{V}' \subseteq \mathcal{V}$ be a non-empty cell of agents and $c \in \mathrm{Fix}^{\bullet}(s)$. Define for all agent $i \in \mathcal{V}'$:*

- *the **deviation from** c as $e_i = x_i - c$.*
- *the **external input** $u_i = \sum_{j \in \mathcal{N}_i \setminus \mathcal{V}'} \frac{a_{ij}}{d_i}(s(x_j) - c)$, with $\|u\|_\infty = \max_{i \in \mathcal{V}'} |u_i|$ being the input magnitude.*
- *the **internal stability coefficient** as $\eta_i = 1 - K d_i^I / d_i$, with $\eta = \min_{i \in \mathcal{V}'} \eta_i > 0$.*

*Moreover, let $\delta = \max_{i \in \mathcal{V}'} |e_i|$ be **the maximum deviation** within the cell. Then, the quantity δ is ISS with respect to the input $\|u\|_\infty$. Specifically, it satisfies for all $t \geq 0$:*

$$\delta(t) \leq e^{-\eta t}\delta(0) + \frac{1}{\eta}(1 - e^{-\eta t}) \sup_{0 \leq \tau \leq t} \|u(\tau)\|_\infty.$$

Consequently, the ultimate bound on the deviation is:

$$\limsup_{t \to \infty} \delta(t) \leq \frac{1}{\eta} \sup_{t \geq 0} \|u(t)\|_\infty.$$

Proof. The Lyapunov-like function we consider is

$$V(\boldsymbol{x}) = \max_{i \in \mathcal{V}'} |x_i - c| = \max_{i \in \mathcal{V}'} |e_i| = \delta.$$

Let us define $m(t) := \arg\max_{i \in \mathcal{V}'} |e_i(t)|$ be the index of the agent with the maximum deviation at time t and denote $e_{m(t)} = x_{m(t)} - c$ its instantaneous deviation.

One has $V(\boldsymbol{x}) = 0$ if and only if $e_m = 0$. Moreover, since V is non differentiable, we use the Dini derivative [129]:

$$\dot{V}(\boldsymbol{x}) = \mathrm{D}^+ V(\boldsymbol{x}(t)) = \limsup_{h \to 0^+} \frac{V(\boldsymbol{x}(t+h)) - V(\boldsymbol{x}(t))}{h}.$$

The Dini derivative $\mathrm{D}^+ V(\boldsymbol{x}(t))$ satisfies $\mathrm{D}^+ V(\boldsymbol{x}(t)) \leq \mathrm{sgn}\,(e_m(t))\, \dot{e}_m(t)$. Thus,

$$\dot{V}(\boldsymbol{x}) \leq \mathrm{sgn}\,(e_m)\, \dot{e}_m. \tag{5.13}$$

Recall that $e_m(t) = x_m(t) - c$. Thus, the dynamics of e_m is

$$\dot{e}_m = \dot{x}_m = \frac{1}{d_m} \sum_{j \in \mathcal{N}_m} a_{mj} s(x_j) - x_m.$$

Let us now separate the sum over neighbors internal to $\mathcal{V}'$ and external to $\mathcal{V}'$, substituting $x_j = c + e_j$ and $x_m = c + e_m$ yields,

$$\begin{aligned}\dot{e}_m &= \frac{1}{d_m}\Big[\sum_{j\in\mathcal{N}_m\cap\mathcal{V}'} a_{mj}s(x_j) + \sum_{j\in\mathcal{N}_m\setminus\mathcal{V}'} a_{mj}s(x_j)\Big] - x_m \\ &= \frac{1}{d_m}\Big[\sum_{j\in\mathcal{N}_m\cap\mathcal{V}'} a_{mj}(s(c+e_j)-c) + \sum_{j\in\mathcal{N}_m\setminus\mathcal{V}'} a_{mj}(s(x_j)-c)\Big] - e_m \\ &= \frac{1}{d_m}\Big[\sum_{j\in\mathcal{N}_m\cap\mathcal{V}'} a_{mj}(s(c+e_j)-c)\Big] + u_m - e_m,\end{aligned}$$

where we used the definition of u_m of the proposition.

Injecting the expression for $\dot{e}_m$ into (5.13), one gets,

$$\dot{V}(\boldsymbol{x}) \leq \operatorname{sgn}(e_m)\frac{1}{d_m}\Big[\sum_{j\in\mathcal{N}_m\cap\mathcal{V}'} a_{mj}(s(c+e_j)-s(c))\Big] - \operatorname{sgn}(e_m)\,e_m + \operatorname{sgn}(e_m)\,u_m$$

Let us upper bound each term in the sum independently.

First, let us notice that from s being Lipschitz continuous around c,

$$\operatorname{sgn}(e_m)\,(s(c+e_j)-s(c)) \leq |s(c+e_j)-s(c)| \leq K|e_j|.$$

Applying this to the sum over neighbors internal to $\mathcal{V}'$, one gets,

$$\operatorname{sgn}(e_m)\Big[\sum_{j\in\mathcal{N}_m\cap\mathcal{V}'} a_{mj}(s(c+e_j)-s(c))\Big] \leq \sum_{j\in\mathcal{N}_m\cap\mathcal{V}'} a_{mj}K|e_j| \leq \sum_{j\in\mathcal{N}_m\cap\mathcal{V}'} a_{mj}KV(\boldsymbol{x}) = d_m^I KV(\boldsymbol{x}).$$

Moreover, we have that $\operatorname{sgn}(e_m)\,e_m = |e_m| = V(\boldsymbol{x})$ and $\operatorname{sgn}(e_m)\,u_m \leq |u_m| \leq \|u\|_\infty$. Thus,

$$\dot{V}(\boldsymbol{x}) \leq \frac{d_m^I}{d_m}KV(\boldsymbol{x}) - V(\boldsymbol{x}) + \|u\|_\infty = -\left(1 - K\frac{d_m^I}{d_m}\right)V(\boldsymbol{x}) + \|u\|_\infty = -\eta_m V(\boldsymbol{x}) + \|u\|_\infty.$$

By assumption, $\eta = \min_{i\in\mathcal{V}'}\eta_i > 0$. Then, we have for all $t \geq 0$:

$$\dot{V}(\boldsymbol{x}(t)) \leq -\eta V(\boldsymbol{x}(t)) + \|u\|_\infty \leq -\eta V(\boldsymbol{x}(t)) + \sup_{0\leq\tau\leq t}\|u(\tau)\|_\infty.$$

By the Gronwall's inequality, this differential inequality implies:

$$V(\boldsymbol{x}(t)) \leq e^{-\eta t}V(\boldsymbol{x}(0)) + \int_0^t e^{-\eta(t-\tau)}\left(\sup_{0\leq\sigma\leq\tau}\|u(\sigma)\|_\infty\right)d\tau.$$

As in the proof of Proposition 5.4.1, a upper bound for the integral term is $\sup_{0\leq\tau\leq t}\|u(\tau)\|_\infty/\eta$. Thus, for all $t \geq 0$:

$$\delta(t) \leq e^{-\eta t}\delta(0) + \frac{1}{\eta}\sup_{0\leq\tau\leq t}\|u(\tau)\|_\infty.$$

Taking the limit as $t \to \infty$, the first term vanishes, and we get:

$$\limsup_{t\to\infty}\delta(t) \leq \frac{1}{\eta}\sup_{t\geq 0}\|u(t)\|_\infty.$$

This completes the proof. □

While Proposition 5.5.1 provides an ISS guarantee for the cell's deviation from a fixed point c under general external influences, it relies on the Lipschitz constant K and does not explicitly consider the specific structure of the signal function $s(x)$ beyond its Lipschitz continuity around c. The following proposition, Proposition 5.5.2, delves deeper into the conditions for ensuring that the states of all agents within a cell $\mathcal{V}'$ remain confined to a specific interval $[c-\varepsilon, c+\varepsilon]$ around the fixed point c. This is a forward invariance property. Instead of a general Lipschitz constant, this result leverages the concept of "consistent estimation" by the signal function s around c, and balances the "restoring force" driving states towards c against the *perturbation amplitude* from external neighbors. This provides a more nuanced understanding of when a cell can maintain its states close to a desired equilibrium, based on the interplay between internal and external degrees and the specific characteristics of the signal function $s(x)$ near c and at the boundaries of the state space.

Proposition 5.5.2. *Let $\mathcal{V}' \subseteq \mathcal{V}$ be a non-empty cell of agents of cardinality N' and c be a stable fixed point of s. i.e., $c \in \mathrm{Fix}^{\bullet}(s)$. For a given $\varepsilon \in (0, \varepsilon_0]$, define the **maximum external perturbation amplitude** of c at ε as:*

$$p(c,\varepsilon) = \max\left(\left|s(1) - (c \pm \varepsilon)\right|, \left|s(-1) - (c \pm \varepsilon)\right|\right),$$

*and the **minimal restoring force amplitude** of c at ε as:*

$$r(c,\varepsilon) = \min\left((c+\varepsilon) - s(c+\varepsilon), s(c-\varepsilon) - (c-\varepsilon)\right).$$

If for a given $\varepsilon \in (0, \varepsilon_0]$, the following condition holds for all $i \in \mathcal{V}'$:

$$d_i^E p(c,\varepsilon) \leq d_i^I r(c,\varepsilon), \tag{5.14}$$

Then, the set

$$\mathcal{H}(c,\varepsilon) = \{\boldsymbol{x} \in \mathcal{X} \mid \text{for all } i \in \mathcal{V}', x_i \in [c-\varepsilon, c+\varepsilon]\}$$

is forward invariant under the dynamics (COSA).

Moreover, such ε exists if either $d_i^E = 0$ or if $s(x_i) \neq x_i$ for all $x_i \in [c-\varepsilon, c+\varepsilon] \setminus \{c\}$ and the following condition holds:

$$\inf_{\varepsilon \in (0,\varepsilon_0]} \frac{p(c,\varepsilon)}{r(c,\varepsilon)} \leq \frac{d_i^I}{d_i^E}$$

for all $i \in \mathcal{V}'$.

Proof. Note that since s is a consistent estimation around c, we have $s(x) \leq x$ for $x \in (c, c+\varepsilon_0]$ and $s(x) \geq x$ for $x \in [c-\varepsilon_0, c)$.

We apply Nagumo's Theorem. For $\mathcal{H}_k(c,\varepsilon)$ to be forward invariant, the vector field $\boldsymbol{f}(\boldsymbol{x})$ must not point outward from $\mathcal{H}_k(c,\varepsilon)$ at any boundary point. The boundary of $\mathcal{H}_k(c,\varepsilon)$ consists of states $\boldsymbol{x}$ where $\boldsymbol{x} \in [c-\varepsilon, c+\varepsilon]^{N'}$, and for at least one $i \in \mathcal{V}'$, either $x_i = c+\varepsilon$ or $x_i = c-\varepsilon$. We need to verify the sign of $\dot{x}_i = f_i(\boldsymbol{x})$ for such boundary nodes i.

Consider $\boldsymbol{x} \in \mathcal{H}_k(c,\varepsilon)$ such that $x_i = c+\varepsilon$ for some $i \in \mathcal{V}'$. We need to show $f_i(\boldsymbol{x}) \leq 0$.

$$f_i(\boldsymbol{x}) = \frac{1}{d_i}\sum_{j \in \mathcal{N}_i} a_{ij} s(x_j) - x_i = \frac{1}{d_i}\sum_{j \in \mathcal{N}_i \cap \mathcal{V}'} a_{ij} s(x_j) + \frac{1}{d_i}\sum_{j \in \mathcal{N}_i \setminus \mathcal{V}'} a_{ij} s(x_j) - (c+\varepsilon)$$

Since $x_j \in [c-\varepsilon, c+\varepsilon]$ for all $j \in \mathcal{V}'$ and s is non-decreasing, we have $s(x_j) \leq s(c+\varepsilon)$ for

$j \in \mathcal{V}' \cap \mathcal{N}_i$.

$$f_i(\boldsymbol{x}) \leq \frac{1}{d_i}\sum_{j\in\mathcal{V}'\cap\mathcal{N}_i} a_{ij}s(c+\varepsilon) + \frac{1}{d_i}\sum_{j\in\mathcal{N}_i\setminus\mathcal{V}'} a_{ij}s(x_j) - (c+\varepsilon) = \frac{d_i^I}{d_i}s(c+\varepsilon) + \frac{1}{d_i}\sum_{j\in\mathcal{N}_i\setminus\mathcal{V}'} a_{ij}s(x_j) - (c+\varepsilon)$$
$$= \frac{d_i^I}{d_i}(s(c+\varepsilon) - (c+\varepsilon)) + \frac{1}{d_i}\sum_{j\in\mathcal{N}_i\setminus\mathcal{V}'} a_{ij}(s(x_j) - (c+\varepsilon)) := \frac{d_i^I}{d_i}(s(c+\varepsilon) - (c+\varepsilon)) + \frac{1}{d_i}p^+_{\text{ext},i}(\boldsymbol{x})$$

By definition of $p(c,\varepsilon)$, we have $|p^+_{\text{ext},i}(\boldsymbol{x})| \leq d_i^E p(c,\varepsilon)$. Therefore, $d_i f_i(\boldsymbol{x}) \leq d_i^I(s(c+\varepsilon) - (c+\varepsilon)) + p^+_{\text{ext},i}(\boldsymbol{x})$.

By the consistent estimation assumption, $s(c+\varepsilon) - (c+\varepsilon) \leq 0$. Thus, the first term is non-positive. For $f_i(\boldsymbol{x}) \leq 0$, it is sufficient that $d_i^I(s(c+\varepsilon) - (c+\varepsilon)) + |p^+_{\text{ext},i}(\boldsymbol{x})| \leq 0$, which requires $|p^+_{\text{ext},i}(\boldsymbol{x})| \leq -d_i^I(s(c+\varepsilon) - (c+\varepsilon))$.

Since $|p^+_{\text{ext},i}(\boldsymbol{x})| \leq d_i^E p(c,\varepsilon)$, the condition $f_i(\boldsymbol{x}) \leq 0$ is guaranteed if $d_i^E p(c,\varepsilon) \leq -d_i^I(s(c+\varepsilon) - (c+\varepsilon))$. This is satisfied by condition (5.14). For the case $x_i = c - \varepsilon$, an analogous argument shows that $f_i(\boldsymbol{x}) \geq 0$ if $d_i^E p(c,\varepsilon) \leq d_i^I(s(c-\varepsilon) - (c-\varepsilon))$.

Since the vector field $\boldsymbol{f}(\boldsymbol{x})$ does not point outward from $\mathcal{H}_k(c,\varepsilon)$ at any boundary point (i.e., $f_i(\boldsymbol{x}) \leq 0$ if $x_i = c+\varepsilon$ and $f_i(\boldsymbol{x}) \geq 0$ if $x_i = c-\varepsilon$), Nagumo's Theorem ensures that the set $\mathcal{H}_k(c,\varepsilon)$ is forward invariant. □

Remark 5.5.1. *The result of Proposition 5.5.2 offers a way to determine an invariant region for the states within cell $\mathcal{C}_k$. If such an interval $[c-\varepsilon, c+\varepsilon]$ is found to be invariant for $\mathcal{C}_k$, and if the overall system states are known to be confined within some global invariant hypercube $[\underline{c}, \overline{c}]^N$ (as established in Proposition 3.6.3), then the external signals $s(x_j)$ for $j \notin \mathcal{C}_k$ are restricted to $[\underline{c}, \overline{c}]$. This could potentially lead to a less conservative estimate for the maximum external perturbation amplitude than $p(c,\varepsilon)$. i.e., $\tilde{\Pi}(c,\varepsilon) = \max(\left(\left|\overline{c} - (c \pm \varepsilon)\right|, \left|\underline{c} - (c \pm \varepsilon)\right|\right), \varepsilon)$, where $\overline{c}$ and $\underline{c}$ are the upper and lower bounds defined in Proposition 3.6.3.*

5.6 Concluding Remarks on Local Synchronization

This chapter provided a multi-faceted analysis of Non-Fully Synchronized Equilibria (NFSE), where agents self-organize into coherent subgroups. This investigation, that is central to understanding phenomena like persistent disagreement, covered two key aspects of local synchronization.

First, we established that *perfect local synchronization*, where agents in a cluster agree exactly, is not a generic outcome but is conditional on the structural symmetry of an equitable partition. This property is significant because it permits an exact model reduction to a low-dimensional quotient system, whose stability cleanly decouples into inter-cluster (quotient) and intra-cluster (transversal) components (Theorem 5.3.11).

Second, for the general case of *approximated local synchronization* in networks lacking perfect symmetry, we used the framework of Input-to-State Stability (ISS). We proved that a cluster's internal disagreement is robustly bounded and that the magnitude of this persistent disagreement is determined not by the total external influence, but by its non-uniform component across the cluster's boundary (Theorem 5.4.1). This result is particularly important because it provides a way to quantify the persistence of disagreement in a network, even in the presence of external influence.

These perspectives are unified: the ISS framework contains the analysis of equitable partitions as the special case where a uniform external influence results in zero residual input,

correctly predicting perfect cohesion. Collectively, the analysis of ideal structures, robust persistence, and dynamic formation provides a comprehensive toolkit for understanding and predicting the emergence of structured disagreement in multi-agent systems with nonlinear interactions.

Chapter 6

Conclusions and Perspectives

6.1 Conclusions

6.1.1 First step of a research journey

This thesis began as an effort to understand how agents, from nations down to individuals, can act to mitigate their impact on climate change. Rather than seeking a single definitive solution, this work was conceived as the first step of a longer research journey. Our approach was to break down this vast problem into three complementary lines of inquiry, each providing a piece of the big picture, but none sufficient on its own. This multi-scale investigation feels all the more urgent as this thesis concludes, at a time when international cooperation on climate remains fragile and contested [131].

6.1.2 Manuscript Conclusion

The aim of this manuscript and results presented was to understand how nonlinear signaling rules and network topology jointly influence the collective behavior in multi-agent opinion dynamics with nonlinear interaction. To make this question formal, we introduced in Chapter 3 and analyzed the *Continuous Opinion Signaled Action* (COSA) model in Chapter 4 and 5, which interpolates between fully transparent and quantized communication while preserving a cooperative structure that enables rigorous stability analysis. The central message is that disagreement or agreement is not solely a property of the graph or of the update rule: it is the outcome of their interaction. Those relationships can be expressed according to spectral terms of the graph and proved with Lyapunov arguments.

At the level of global behavior in Chapter 4, the thesis established a threshold for network-wide synchronization $\lambda_{N-1}K < 1$, where K is the maximal nonlinear gain of the signal and λ_{N-1} that measures the cohesion of the graph. This condition is sharp in the sense that: it is a sufficient condition for global synchronization and its violation is sufficient to construct a signal function s that induces a persistent disagreement. Moreover, we explicited some necessary conditions that the signal function must verify to induce a persistent disagreement that are solely based on the shape of the signal function.

When global agreement is not enforced, the analysis in Chapter 5 characterized the structure of persistent disagreement looking at two different aspects: *perfect* local synchronization and *approximated* local synchronization. Exact cluster agreement requires structural symmetries

[131] "Missing at U.N.'s Climate Meeting". Ivan Penn et al. (2025).

and is captured by regular and equitable partitions on some quotient dynamics. In generic networks that lack perfect symmetry, the thesis used input-to-state stability to bound within-cluster dispersion under external influence, formalizing *approximated* local synchronization in terms of interpretable gains and perturbation levels.

Methodologically, the manuscript advances a way of reasoning that separates *mechanism* (the signal function) from *structure* (the graph) and studies their coupling in a geometry aligned with disagreement modes. Conceptually, the results clarify that cooperative networks with nonlinear signaling can display clustered equilibria without invoking antagonistic edges. In reality, what matters is whether the signal amplifies differences faster than the topology can dissipate them. Practically, the conditions derived here act as guidelines: they tell a designer when a given combination of network and signaling law will synchronize, when clusters will remain coherent.

6.2 Perspectives

As usual in research this research journey opens more question than it is closing. The following is a survey of more or less direct extension and question that could and should be answered. We organize the perspective per research stage.

6.2.1 First Stage Perspectives: Designing Games for Cooperation

A central difficulty in environmental game theory is that many models suggest robust, large-scale cooperation won't occur unless there is external enforcement. A path forward does not lie in building ever-more-complex models that simply confirm this difficulty numerically, turning them into analytically opaque "black boxes" and falling back into the IAMs traps. Instead, the objective should be to strategically enrich the models to explicitly search for and analyze **mechanisms that can facilitate agreement**. The aim is to move from diagnosing the failure of cooperation to understanding how to design its success. With this goal in mind, the following research directions become particularly relevant.

Designing Robust and Incentive-Compatible Policy Mechanisms. The direct way to promote cooperation is to revise the game's rules. Future work should design and analyze concrete policy tools that ensure ambitious climate action is incentive-compatible. This involves moving beyond the basic emissions game to incorporate structures from **contract theory** and **mechanism design**. For example, one could formalize and test the stability of treaties that include systems of credible sanctions, such as *fees for deviating* from an agreed emissions trajectory. Another line of research is to model a *global carbon price* where revenues are recycled into a common investment fund for green technology, creating a positive feedback loop that rewards participants. Analyzing such mechanisms can provide concrete insights into how the architecture of an international agreement can fundamentally shift strategic incentives from free-riding towards mutual reinforcement.

Leveraging Social Dynamics and Endogenous Preferences. The current framework assumes static preferences (as the utility function is fixed), yet a nation's political interests are dynamic. A important extension is to endogenize preferences by coupling the strategic game with models of *opinion dynamics and social influence*. If an agent's perceived damage from climate change, or

its tendency to imitate others, changes based on its neighbors' actions and signals, the game's payoffs are no longer fixed. This opens the possibility of identifying leverage points. For instance, a "meta-game" of *lobbying* could be modeled, where influential nations strategically target others to shift collective norms towards greater climate ambition. This research area connects rational choice theory, that is game theory, and behavioral dynamics, that is opinion dynamics, aiming to understand how social and political influences can be shaped to promote cooperation. It could also incorporate agent heterogeneity, analyzing how a critical mass of *aware agents* can influence, or not, the behavior of more *myopic* ones.

Grounding Strategy in Realistic Economic and Geophysical Dynamics. To ensure that proposed mechanisms are viable (in a strategic point of view), they must be build upon realistic representations of the economy and the environment. Integrating a more sophisticated macroeconomic structure, such as a *neoclassical growth model* (e.g., Solow-Ramsey or less optimalistic but still tractable models) would frame decarbonization in terms of the fundamental trade-offs between investment in capital and consumption. This allows for an *optimal control* perspective, identifying efficient decarbonization pathways that are also aware of economic dynamics. Similarly, introducing *stochastic events*, such as climate tipping points, is essential for testing the robustness of any proposed agreement. The goal of this integration is again not to create a complex simulation for the sake of realism. Instead, it is to ensure that the strategic analysis is informed by real-world constraints and uncertainties. A model that is easy to analyze, even with enhanced features, can explain how shifting economic dynamics and high risks change the strategic situation and may make long-term cooperation more attractive.

6.2.2 Second stage perspectives

The (COSA) framework provides a foundation for analyzing the interplay between networked agents and a responsive environment. The next stage of research involves extending this framework to capture more complex, system-level phenomena and structural realities. These extensions move beyond the analysis of a single, homogeneous system to explore heterogeneity, multi-scale interactions, and more sophisticated environmental dynamics. Here are some of the directions that could be explored:

Community Structure and Heterogeneous Parameters. Real-world networks are rarely homogeneous; they are typically organized into communities or clusters. This structure can be explicitly modeled by assuming that parameters like the environmental trust, β, are community-dependent (β_k for community k). Furthermore, communities might interact with distinct local environments (e_k) that are weakly coupled. This leads to a model with heterogeneous dynamics structured by the underlying graph topology. The analysis would focus on the stability of clustered states, where consensus is achieved within communities but disagreement persists between them. Important research questions include identifying conditions under which inter-community influence achieves global consensus versus those that disagreement persists between them, especially as local environmental feedback may induce sustained oscillations within each community.

Multi-Environment Interactions. Agents are often influenced by multiple, distinct environmental signals simultaneously. For instance, public opinion on a policy might be shaped by its perceived environmental impact (e_1), economic consequences (e_2), and public health outcomes

(e_3). This requires extending the environmental state to a vector $\boldsymbol{e} \in \mathbb{R}^m$. The agents' opinion dynamics would then be driven by a function of this vector, e.g.,

$$\dot{\boldsymbol{x}} = (1-\beta)\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) + \beta\boldsymbol{r}(\boldsymbol{e}), \tag{6.1}$$

where $\boldsymbol{r} : \mathbb{R}^m \to [-1,1]^N$ is a vector of response functions representing the relative importance of each environmental signal. This formulation introduces the possibility of trade-offs and more complex equilibria and represent a direct extension of [69]. Polarization could occur not just on a single opinion axis, but in a multi-dimensional belief space, as different subgroups prioritize different environmental factors. This could be formalized by introducing a vector of weights $\boldsymbol{w} \in [0,1]^m$ for each environmental signal, such that $\boldsymbol{w}^\top \mathbf{1} = 1$. This work would extend the results from [132], where a simpler model of internal state feedback was considered. Moreover, we could extend the model incorporating influence between the opinions on each environment (e.g., economic opinion could influence the environmental opinion).

Nonlinear and Multistable Environmental Dynamics. The assumption of a linear, first-order environmental dynamic, as currently considered, is a simplification. Many real-world processes, such as epidemics (SIS/SIR models), ecological systems, or economic markets, are governed by nonlinear dynamics with multiple equilibria, tipping points, or intrinsic oscillations. Coupling the opinion dynamics model with such an environment creates a far richer feedback system. For example, if the environment is a bistable system, the collective opinion of the agents could act as a control input that pushes the environment across a separatrix, causing an irreversible shift [133]. The analysis would necessarily rely on bifurcation theory to understand how the parameters of the opinion model (e.g., β) interact with the intrinsic bifurcations of the environmental model, potentially creating new, complex system-level behaviors.

Networks of Coupled (COSA) systems. The current model analyzes a single population coupled to one environmental state. A important generalization is to consider a network of interconnected (COSA) systems, where each node represents an entire community (e.g., a city, a country) where the agregated opinion of connected communities act as environment. The state space becomes the product of the individual systems' state spaces, and the coupling introduces a new topological layer. Bifurcation analysis, as developed in works like [69], becomes a promising tool for understanding how inter-system coupling can induce collective phenomena, such as synchronized oscillations or cascading bifurcations. This abstraction has a profound impact beyond social dynamics, offering a powerful model for computational neuroscience. In this context, each system represents a neural population, its aggregated opinion mirrors the local field potential, and the inter-system network maps to the brain's connectivity. The framework can thus be used to analyze the emergence of brain rhythms, inter-regional coherence, and how pathological bifurcations could model the onset of neurological disorders like epileptic seizures. This would extend results from Chapter 4 and Chapter 5 to a coupled network setting with broad applicability.

Robustness and Input-to-State Stability (ISS) of Disagreement. Instead of modeling the environment as a deterministic state, it can be viewed as an exogenous, unknown-but-bounded

[69] *Bifurcation Analysis of an Opinion Dynamics Model Coupled with an Environmental Dynamics*. Anthony Couthures et al. (2025).
[133] "Opinion-Driven Risk Perception and Reaction in SIS Epidemics". Marcela Ordorica Arango et al. (2025).

disturbance input. This deviates the analytical goal from characterizing equilibria to guaranteeing robustness. The primary concern is the persistence of polarized clusters despite such environmental perturbations. This can be formalized using the tools of robust control, particularly the ISS framework as in Chapter 5. The objective would be to prove that the disagreement between clusters is ISS with respect to the environmental signal, and that whatever the environmental state. A formal ISS guarantee would ensure that if environmental trajectories are bounded, the deviation from a perfectly polarized state is also bounded, providing a rigorous measure of the resilience of social structures like echo chambers, as a preliminary result was explored in the context of the (CODA) model [66].

Stochastic and Event-Driven Environmental Feedback. Environmental feedback is often not continuous in time often feedback is recieved at discrete times events. Moreover, the time at wich the feedback is recieved is often not periodic and likely from a random process: corresponding to discrete events like natural disasters, policy announcements, or market shocks. Modeling this requires a change in formalism from ordinary differential equations to either Stochastic Differential Equations (SDEs) or hybrid/impulsive systems. In the event-driven case, the framework of hybrid systems could be a proper tool to analyze the dynamics where continuous evolution is punctuated by discrete resets or changes in the system parameters, which is a more realistic model for the types of information provided by bodies like the IPCC [5] . Moreover, the feedback could be a combination of continuous and discrete components, such as a continuous feedback with a discrete component, or a discrete feedback with a continuous component.

6.3 Manuscript perspectives

6.3.1 Extend the nonlinear interaction model

The current (COSA) model is built upon several key assumptions that ensure analytical tractability while capturing the essential dynamics of signaled consensus. These assumptions, however, represent simplifications of more complex real-world scenarios and theoretical framework. Relaxing them opens up numerous directions for future research, each with its own set of modeling opportunities and mathematical challenges. Here, we briefly discuss some of the most promising extensions.

Individual Signal Functions (s_i). Our current framework assumes a common signal function s for all agents, representing a population with homogeneous interactions. A natural extension, but far more complex, is to consider individual signal functions, $s_i : [-1, 1] \to [-1, 1]$, for each agent $i \in \mathcal{V}$. This would allow for modeling heterogeneous systems where agents have different communication styles, cognitive biases, or technological capabilities (e.g., some agents might be under-estimators, others over-estimators). From a mathematical perspective, this would alter the dynamics on the synchronization manifold and will not be forward-invariant for the dynamics anymore. A synchronized equilibrium $\boldsymbol{x}^* = c\mathbf{1}$ would no longer be a fixed point of a single function, but would instead need to satisfy a more complex self-consistency condition, such as $c = \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} s_j(c)$ for all i, which is far more restrictive. The study of equilibria would

[66] "Analysis of a Continuous Opinion and Discrete Action Model Coupled with an External Dynamics". Anthony Couthures et al. (2024).

[5] As seen in the IPCC Atlas of Climate Change https://interactive-atlas.ipcc.ch/

shift from analyzing a scalar function to solving a system of nonlinear equations. This analysis would follow the work of [126] where their framework has those individual signal functions.

Directed, Weighted, and Antagonistic or Competitive Interactions. We have assumed the communication graph $\mathcal{G}$ to be undirected and unweighted with non-negative influence. Relaxing this opens several avenues:

– **Directed graphs** would allow for modeling asymmetric influence, such as in social networks with leaders and followers or in hierarchical structures. The main implication of moving to non-symmetric adjacency matrices is the loss of the clean spectral theory associated with symmetric operators. In this case, eigenvalues can become complex even if the system is still cooperative, thus we can expect that almost all trajectory will converge to equilibria. More over, the application to flow graphs would be interesting to explore as the directed communication can be seen as an endogenous influence over the opinion of the agents. This asks multiple question as, what would be the role of the left eigenvector of the row-stochastic matrix? Will it be the same as in the linear model since it is in this case it does not induce an invariant for the dynamics?

– **Antagonistic interactions: symmetric interaction with negative weights** would model contrarian behavior, mistrust, or what one can call "agree to disagree" behavior. One can assume that allowing $a_{ij} < 0$ could fondamentaly alter the cooperative structure of the system but it will likely not be the case if the graph is symmetric. Even if naively, assuming a smooth signal function s, would lead to a Jacobian matrix that is not Metzler, and thus forbid the usage of results of monotone systems theory, the system could still be $\mathcal{K}$-cooperative (where $\mathcal{K}$ is an orthant of $\mathbb{R}^N$) [126, 134]. For the linear model (Laplacian Model) with antagonistic interaction on unbalanced graph, the expected behavior is that opinion will collapse to $\mathbf{0}$. However, in the presence of the nonlinear interaction function s, if the nonlinearity sufficiently amplify the opinion, polarization would still emerge even in this unbalanced case. Providing the insight of Chapter 4 and Chapter 5 would allows to better understand the dynamics moving beyond mainly a bifurcation analysis to characterize the persistent polarization and disagreement in this context.

– **Competitive interactions: asymmetric interaction with negative weights** would model competitive behavior, such as in a prey-predator model where the "pursuer-evader" behavior would likely emerge. This could be modeled by allowing $a_{ij} < 0$ and $a_{ji} > 0$ for some $i, j \in \mathcal{V}$ with $i \neq j$. This opens the door to far more complex behaviors, including sustained oscillations (limit cycles) and potentially chaos, requiring a shift to tools like bifurcation analysis or contraction-based methods. This analysis could complement the understanding of [134] where their framework has only symmetric antagonistic interactions using insight from Papers A.2 and A.3.

Dynamic Topologies. Our model assumes a static graph. In many real-world applications, from mobile robotic networks to evolving social structures, interactions are dynamic. This can be modeled in two distinct ways:

– **Time-Varying Topologies and Asynchrony:** A direct extension for the discrete time model (COSA-DT) is to consider an exogenously changing network topology, which renders the system non-autonomous and described by the linear update rule $\boldsymbol{x}(t+1) =$

[126] "Multiequilibria Analysis for a Class of Collective Decision-Making". Angela Fontan et al. (2018).
[134] "The Role of Frustration in Collective Decision-Making Dynamical Processes". Angela Fontan et al. (2022).

$\boldsymbol{A}(t)\boldsymbol{x}(t)$. The central analytical challenge becomes understanding the convergence properties of the infinite product of row-stochastic matrices. A foundational requirement for consensus is that the graph of interactions remains connected over time [78]. Formally, the union of the communication graphs over any infinite time horizon must be strongly connected. However, as demonstrated in that literature, this connectivity condition alone is not sufficient to guarantee consensus for general directed graphs. An additional condition on the communication frequency, such as the "bounded intercommunication intervals" assumption, is required to rule out pathological update sequences that can prevent convergence. It is only under specific structural assumptions, such as communication symmetry, that the connectivity condition alone becomes sufficient.

– **State-Dependent (Adaptive) Graphs:** A more complex and realistic scenario is when the topology is endogenous, co-evolving with the agents' states, i.e., $\boldsymbol{A}(\boldsymbol{x}(t))$. For example, agents might cut or change the nature (from cooperative to antagonistic), of the connections to those with very different views. This change in the model could illustrate how echo chambers arise, as numerically observed in [98, 100]. Additionally, a state dependent graph introduces a strong feedback loop where opinions influence the network dynamics and the network dynamics influences opinions. Mathematically, this leads to a highly nonlinear autonomous system where the vector field itself depends on the state in a potentially discontinuous manner (e.g., if $a_{ij}(\boldsymbol{x})$ is a step function). Such systems can lead to network fragmentation and require sophisticated tools from switched systems or non-smooth analysis.

Discontinuous Signal Functions. While our standing assumption requires Lipschitz continuity for the well-posedness of the ODE system, many practical scenarios, including the original CODA model, are best described by discontinuous signal functions (e.g., step functions for quantization) [105]. Incorporating such discontinuities changes the mathematical nature of the system. The dynamics would no longer be described by an ordinary differential equation, but rather by a *differential inclusion*, with solutions understood in the sense of Filippov or Krasovskii [104]. This framework is necessary to handle the ambiguity of the vector field at points of discontinuity. It opens the door to richer dynamical behaviors, such as sliding modes along disagreement surfaces, but requires a more advanced, non-smooth analytical toolbox.

Multi-Dimensional Agent States. The (COSA) model assumes each agent holds a single scalar opinion x_i. A important extension would be to consider vector states $\boldsymbol{x}_i \in \mathbb{R}^{m_i}$, where agents have opinions on multiple interconnected topics or physical states (e.g., position and velocity) as in [96, 97]. Communication could be partial, where an agent's signal $\boldsymbol{s}_i(\boldsymbol{x}_i) \in \mathbb{R}^k$ with $k < m$ reveals only a subset of its internal state. This would allow for modeling complex phenomena where consensus on one topic might influence disagreement on another. The analysis would move from scalar dynamics to a high-dimensional system of coupled equations, requiring tools from matrix theory and the study of cooperative systems in higher dimensions.

[78] "Convergence in Multiagent Coordination, Consensus, and Flocking". V.D. Blondel et al. (2005).
[98] "Modeling Echo Chambers and Polarization Dynamics in Social Networks". Fabian Baumann et al. (2020).
[100] "Transitions between Polarization and Radicalization in a Echo-Chamber". Łukasz G. Gajewski et al. (2022).
[105] "Consensus and Disagreement". Francesca Ceragioli et al. (2018).
[104] "Discontinuities, Generalized Solutions, and (Dis)Agreement in Opinion Dynamics". F. Ceragioli et al. (2018).
[96] "Nonlinear Opinion Dynamics With Tunable Sensitivity". Anastasia Bizyaeva et al. (2023).
[97] "Multi-Topic Belief Formation Through Bifurcations". Anastasia Bizyaeva et al. (2025).

Intrinsic Agent Dynamics. The current model assumes agents are passive responders to social and environmental signals. An important extension is to allow agents with intrinsic dynamics, representing internal psychological processes or biases. This could be modeled by adding a term $h_i(x_i)$ to the opinion update rule. If $h_i(x_i)$ represents a drift towards a prejudice, it introduces stubbornness. More complex forms, such as second-order internal dynamics, could model agents as intrinsic oscillators. The problem then becomes one of analyzing a network of coupled oscillators subjected to a common environmental forcing signal. This connects the research to the rich field of synchronization theory, including phenomena like phase-locking, frequency entrainment, and oscillator death, where the environmental feedback could either promote or suppress synchronous behavior.

Analysis through Contraction and Dissipativity Theory. In the context of directed or competitive interactions the system's most interesting behaviors, such as limit cycles and chaos, occur away from simple equilibria. While Lyapunov theory is effective for local stability analysis of fixed points, a more global understanding of these complex dynamics requires a different set of tools. Contraction and dissipativity theory offer a promising direction.

- **Contraction analysis** guarantees that all system trajectories converge to a single, potentially time-varying, trajectory if the system's Jacobian (under smoothness assumptions) has a uniformly negative matrix measure, provides a powerful non-local proof for the stability of attractors like limit cycles [135]. This could use a even more general notion of contraction, such as k-contraction, to analyze the complexity of the system's dynamics.
- **Dissipativity theory** offers a compositional, input-output perspective by analyzing the system as a feedback interconnection of the opinion and environmental dynamics. Proving that each subsystem dissipates a form of "energy" can guarantee the stability of the entire coupled system, providing a method for analyzing the boundedness of trajectories in the chaotic regime [136].

6.3.2 Dynamics of Polarization and Recruitment

While Chapter 5 characterized the *stability* of clusters via the ISS framework, the dynamic processes of cluster *formation* and *expansion* remain open. We propose that the interplay between local graph structure and the signal function s dictates not just the static existence of clusters, but the dynamic "recruitment" of agents. The following are currently investigated topics:

Formalizing Robust Polarized Clusters (RPC). A natural next step is to rigorously define a "Robust Polarized Cluster" as a set of agents capable of maintaining a specific opinion range (e.g., strictly above an unstable threshold c°) against arbitrary external signals. Unlike the ISS analysis which bounds deviation from a mean, this would focus on the *invariance* of specific regions in the state space. We conjecture that a "tug-of-war" condition exists, comparing the internal "pull" of a cluster to the external "resistance," which would serve as a sufficient condition for such clusters to persist.

Recruitment and Propagation Cascades. Understanding how polarization spreads is as critical as understanding how it persists. Future research should investigate the mechanism of

[135] *Contraction Theory for Dynamical Systems*. F. Bullo. (2024).
[136] *Dissipative Systems Analysis and Control*. Bernard Brogliato et al. (2007).

"recruitment," where an established cluster pulls a neighboring boundary agent across an opinion tipping point. We hypothesize that this local recruitment process can trigger network-wide **cascades**, where the recruitment of one layer of agents changes the neighbor ratio for the next layer, leading to a domino effect. Quantifying the speed of these cascades and identifying the "stopping sets" that halt them would be a contribution to understanding radicalization and polarization dynamics.

Geometric Constraints and the Cheeger Constant. We further conjecture that the "shape" of a cluster plays a critical role in its ability to recruit neighbors and maintain polarization. Intuitively, the nonlinear "holding force" of the signal function must exceed the "leakage" of influence across the cluster boundary. This suggests a rigorous link between the *Cheeger constant* (conductance) of a subgraph and the nonlinear gain of the system. Establishing this link would allow us to predict polarization based purely on spectral properties (like λ_2), potentially proving that "expander graphs" (high connectivity) are immune to polarization regardless of the signal function's steepness.

6.3.3 Community structure and a "new" approach for clustering

The emergence of polarized states, as analyzed in Chapter 5, is rooted linked to the community structure of the underlying communication network. This connection suggests that the theoretical tools developed in this thesis, particularly ISS theory, could offer a new, dynamics-centric perspective on community detection.

A Spectral Definition of Community Robustness. The ISS framework allows us to move beyond static, count-based definitions of communities. A interesting direction of research is to formalize a dynamically robust community using the spectral properties of its induced subgraph. The ISS condition from Chapter 5 can be translated into the stability margin $\alpha_{\text{in}} = 1 - K \max_i |\lambda_i(\boldsymbol{D}_{\text{in}}^{-1}\boldsymbol{A}_{\text{in}})| > 0$. This raises the question: how does this definition of a "good" community, based on its ability to withstand nonlinear dynamic perturbations, compare to classical notions based on modularity or conductance? And can this spectral condition serve as a more natural measure of network cohesion? Benchmarks have to be build to compare the approaches present in the literature and their performances.

Exploring Dynamics-Based Clustering Paradigms. This perspective opens the door to a new class of community detection algorithms not based on optimizing edge counts, but on maximizing dynamical resilience. For instance, one could explore heuristic algorithms guided by the principle of maximizing the minimal stability margin α_{in} across a partition. This poses both theoretical and practical questions:

- **Algorithmic Design:** How can one efficiently approximate the solution to maximizing the minimum stability margin?
- **Performance:** Would such an approach uncover meaningful structures in networks where traditional methods fail, such as in functional brain networks or sparse, irregular social graphs where the notion of "density" is less informative?

Relating Stability-Dependent and Rate-Dependent Clustering. The decaying confidence model of [88] also forms communities, but through a mechanism dependent on convergence

[88] "Opinion Dynamics With Decaying Confidence". Irinel-Constantin Morarescu et al. (2011).

rates. A compelling line of investigation would be to formally understand the relationship between these two mechanisms. Do rate-dependent and stability-dependent clustering capture the same structural features, or do they reveal complementary aspects of network organization: one tied to the speed of consensus, the other to the resilience of disagreement? A first insight was present in Figure 5.5 where the communities seems to be the same for the two approaches. However, the theoritical foundation of this observation is still missing.

Communities in the Presence of Antagonism. Extending the notion of a robust community to antagonistic networks presents a significant challenge, as the concept of cohesion must be reconciled with repulsive forces. This leads to several open problems:

- **Structurally Balanced Clusters:** Can a structurally balanced subgraph be proven to be ISS with respect to external influence, thus formalizing its role as a stable "antagonistic community"?
- **Frustrated Clusters**[6] **:** The more complex case of frustrated subgraphs, which may lack a stable internal equilibrium, remains largely unexplored. Do they act as inherently fragile points in the network, or can they be stabilized by their surrounding structure? Answering this would require extending the ISS tools beyond the cooperative setting.

From Network Analysis to Network Generation. This framework suggests a paradigm shift from network analysis to network generation. Rather than asking what communities a given graph possesses, one could ask: for a given interaction law s, what network topologies will support a desired partition into robust communities? This involves designing a graph where two conditions are simultaneously met:

- **Intra-community stability conditions** are satisfied, ensuring each cluster's internal cohesion.
- **Inter-community propagation conditions** are deliberately violated, preventing recruitment and ensuring the partition's stability.

This presents a potential direction for engineering resilience or modularity in distributed systems.

A Multi-Scale View via Timescale Separation. In large networks, a natural timescale separation often exists between fast intra-community dynamics and slow inter-community influence. A promising direction is to exploit this using singular perturbation theory. This could allow for a formal model reduction, where the fast dynamics within each robust community are abstracted away, and the slow dynamics are described by a reduced-order model of interacting "super-nodes" [137]. This would not only offer a tractable approach for massive networks but would also formalize the intuitive notion of communities as coherent, higher-level agents. Another approach that would need to be compared to the rate-dependent clustering approach.

[6] In the sense of [134] that relate with the presence of negatives edges inside a dense cluster.

[137] "Novel Results on Slow Coherency in Consensus and Power Networks". Diego Romeres et al. (2013).

Appendix A

Paper of the second stage: Interplay of Environment and Opinion Dynamics

Analysis of a continuous opinion and discrete action dynamics model coupled with an external observation dynamics*

A. Couthures[1], T. Mongaillard[1], V.S. Varma[1,2], S. Lasaulce[3,1], I.C. Morărescu[1,2]

***Abstract*— We consider a set of consumers in a city or town (who thus generate pollution) whose opinion is governed by a continuous opinion and discrete action (CODA) dynamics model. This dynamics is coupled with an observation signal dynamics, which defines the information the consumers have access to regarding the common pollution. We show that the external observation signal has a significant impact on the asymptotic behavior of the CODA model. When the coupling is strong, it induces either a chaotic behavior or convergence towards a limit cycle. When the coupling is weak, a more classical behavior characterized by local agreements in polarized clusters is observed. In both cases, conditions under which clusters of consumers don't change their actions are provided.Numerical examples are provided to illustrate the derived analytical results.**

I. Introduction

Opinion dynamics (OD) over social networks attracted a lot of attention during the last decades. Multi-agent systems have provided an efficient way to model opinion evolution under social interactions. The existing OD models consider that the opinions evolve either in a discrete set [1], [2], [3], [4] or in a continuous set of values [5], [6], [7], [8]. While some models naturally lead to consensus [9], [7] some others yield a network clustering [5], [6], [8], [10]. However, all the models enumerated above consider that each individual has access to the opinion values of the neighbors. In order to more accurately describe the opinion dynamics and to recover more realistic behaviors, a mix of continuous opinion with discrete actions (CODA) was proposed in [11]. This model reflects the fact that even if we often face binary choices or actions that are visible to our neighbors, our opinion evolves in a continuous space of values that are not accessible. A consensus-like dynamics reproducing this behavior has been proposed and analyzed in [12] where the preservation and the propagation of actions are also characterized through the notion of robust polarized clusters. While the model in [12] led to a clustering of the network, a similar idea was employed in [13] to study the emergence of consensus under quantized all-to-all communication.

In this paper we analyze the behavior of the CODA model introduced in [12] coupled with an external dynamics. Many models have been developed to characterize the pollution dynamic in urban areas, considering the fluid dynamics approach [14], chemistry-based approach [15], or both [16]. Even if the time constants depend on the chemical compound considered [17], we introduce a simple linear pollution model to estimate the local air quality. In this model, the pollution level depends on the actions of the individuals which in turn are influenced both by the actions of their neighbors and the pollution level. The coupling of the two dynamics leads to a complex asymptotic behavior that can be summarized as follows. When the coupling between the dynamics is weak, one recovers the asymptotic behavior of the original CODA model in [12]. A strong coupling between the two dynamics hampers the convergence towards a steady state and yields either chaotic oscillations or convergence towards a limit cycle. It is noteworthy that even in the simplified case when all the agents have the same initial opinion, the strong coupling with the external dynamics hampers the convergence toward a steady state and may lead to chaotic oscillations.

The main contributions of this paper are: i) the introduction of a mathematical model capturing the coupling between the CODA dynamics and an external one; ii) the analysis of the asymptotic behavior of the aforementioned model; iii) and the characterization of the coupling strength leading to different asymptotic behaviors.

The paper is structured as follows. Section II presents the definitions of the measures that constitute the model. Characteristics of opinion equilibrium and asymptotic behavior are analyzed in Section III, followed by a focus on the synchronized behavior in Section IV. Section V illustrates the different behaviors with numerical simulations. Finally, Section VI concludes our work.

II. Problem formulation and preliminaries

Model description

We consider the classical multi-agent modeling in which n individuals/agents belong to the set $\mathcal{V} = \{1, \dots, N\}$ and interact according to a fixed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that can be directed or not. The neighborhood of the agent i is denoted by $\mathcal{N}_i$ and represents the set of agents that influence i according to the graph $\mathcal{G}$ (i.e $j \in \mathcal{N}_i \Leftrightarrow (j,i) \in \mathcal{E}$). We also denote by n_i the cardinality of $\mathcal{N}_i$. We assign to each agent $i \in \mathcal{V}$ an opinion $\theta_i \in [-1, 1]$ that evolves in time according to a discrete time protocol defined further in (3). Let $\theta_i(k)$ be the opinion of the agent $i \in \mathcal{V}$ at time k and $\theta(k)$ the opinion of all individuals at time k. Let us also introduce the action value $q_i(k) \in \{-1, 1\}$ as a quantized

*This work was partially supported by ANR through the grant NICETWEET No. ANR-20-CE48-0009 and the PNRR project DECIDE No. 760069.

[1] Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France. anthony.couthures@univ-lorraine.fr

[2] associated with Automation Department, Technical University of Cluj-Napoca, Memorandumului 28, 400114 Cluj-Napoca, Romania.

[3] associated with Khalifa University, Abu Dhabi, UAE.

version of $\theta_i(k)$ defined by

$$q_i(k) = \begin{cases} 1 & \text{if } (\theta_i(k) > 0) \vee (\theta_i(k) = 0 \wedge q_i(k-1) = 1) \\ -1 & \text{if } (\theta_i(k) < 0) \vee (\theta_i(k) = 0 \wedge q_i(k-1) = -1) \end{cases}$$

We assume that the action of individual i at time k generates an emission $e_i(k) \in [e_{\min}, e_{\max}] \subset \mathbb{R}$ where $e_{\min}$ and $e_{\max}$ are the minimum and maximum emissions, respectively. The emission is given by the following equation

$$e_i(k) = \begin{cases} e_{\min} \text{ if } q_i(k) = -1 \\ e_{\max} \text{ if } q_i(k) = 1 \end{cases}. \tag{1}$$

We add an external state $p \in \mathbb{R}$, referred to as pollution, that captures the environment state under the emission of everyone. The pollution evolves according to the following discrete-time dynamics:

$$p(k+1) = \gamma p(k) + \sum_{i=1}^{N} e_i(k), \tag{2}$$

where $\gamma \in (0,1)$ is an autonomous decay rate.

We assume that individuals cannot observe $p(k)$ but they can sense a quantized value $q_p(k) \in \{-1, 1\}$. Let us define $q_p(k)$ as a function of a threshold $\bar{p} \in \mathbb{R}$ as follows:

$$q_p(k) = \begin{cases} -1 \text{ if } (p(k) > \bar{p}) \vee (p(k) = \bar{p} \wedge q_p(k-1) = -1) \\ 1 \quad \text{if } (p(k) < \bar{p}) \vee (p(k) = \bar{p} \wedge q_p(k-1) = 1). \end{cases}$$

We are now ready to describe the opinion dynamics model that we consider in this work. This dynamics adapts the CODA model in [12] to include the external dynamics of $p(k)$:

$$\begin{aligned} \theta_i(k+1) = \theta_i(k) + \left(1 - \theta_i(k)^2\right) \big[& \beta \left(q_p(k) - \theta_i(k)\right) \\ & + (1-\beta) \frac{1}{n_i} \sum_{j \in \mathcal{N}_i} \left(q_j(k) - \theta_i(k)\right) \big], \end{aligned} \tag{3}$$

where $0 \leq \beta \leq 1$ encapsulates the trade-off between the environment state observed through $q_p(k)$ and the opinions of the neighbors. We note that the complete model coupling CODA and the external dynamics is described by (1)-(3).

We emphasize a natural partition of $\mathcal{V}$ in two subsets $\mathcal{N}^-(k) = \{i \in \mathcal{V} \mid q_i(k) = -1\}$ and $\mathcal{N}^+(k) = \{i \in \mathcal{V} \mid q_i(k) = 1\}$. In the following, we denote by $n^-(k)$ and $n^+(k)$ the cardinality of $\mathcal{N}^-(k)$ and $\mathcal{N}^+(k)$, respectively. Similarly, for an agent i we denote by $\mathcal{N}_i^-(k) = \mathcal{N}_i \cap \mathcal{N}^-(k)$ and $\mathcal{N}_i^+(k) = \mathcal{N}_i \cap \mathcal{N}^+(k)$ and by n_i^+ and n_i^- the cardinalities of these sets.

III. Analysis of the model

Before starting the analysis of the model introduced in the previous section, let us observe that extreme opinion values $\theta_i(0) \in \{-1, 1\}$ do not evolve in time. We also observe that the definitions of q and q_p are rigorous only if $\theta_i(0) \neq 0, \forall i \in \mathcal{V}$ and $p(0) \neq \bar{p}$. Therefore, the following assumption is perfectly justified by our setup.

Standing Assumption 1. *For all $i \in \mathcal{V}$, $\theta_i(0) \in (-1,1)\backslash\{0\}$ and $p(0) \neq \bar{p}$.*

A. Characterization of opinion equilibria

In the following, we analyze the asymptotic behavior of opinions that follows the dynamics (3). In other words, we assume that the external signal has an exogenous decoupled evolution.

To simplify our further reasoning, we introduce the following notation

$$f_i(k) = (1-\beta) \frac{n_i^+(k) - n_i^-(k)}{n_i} + \beta q_p(k). \tag{4}$$

Lemma 1. *Let $i \in \mathcal{V}$, $\theta_i(0) \in (-1,1)$. Then for all $k \in \mathbb{N}$, one of the following relation holds*

$$\theta_i(k) < \theta_i(k+1) < f_i(k), \tag{5}$$

$$\theta_i(k) > \theta_i(k+1) > f_i(k), \tag{6}$$

or,

$$\theta_i(k) = \theta_i(k+1) = f_i(k). \tag{7}$$

Proof. Let us first observe that $\sum_{j \in \mathcal{N}_i} q_j(k) = n_i^+(k) - n_i^-(k) = 2n_i^+(k) - n_i$, since $n_i^+(k) + n_i^-(k) = n_i$ for all $k \in \mathbb{N}$. Then, using (4), one rewrites (3) as:

$$\begin{aligned} \theta_i(k+1) &= \theta_i(k) + \left(1 - \theta_i(k)^2\right) \big[\beta \left(q_p(k) - \theta_i(k)\right) \\ &\qquad + (1-\beta) \frac{1}{n_i} \left(n_i^+(k) - n_i^-(k) - n_i \theta_i(k)\right) \big] \\ &= \theta_i(k) + \left(1 - \theta_i(k)^2\right) \left[f_i(k) - \theta_i(k)\right] \end{aligned} \tag{8}$$

We continue our reasoning by induction. From equation (8) it is straightforward that if $\theta_t(k) < f_i(k)$ then $\theta_i(k) < \theta_i(k+1)$ and $\theta_i(k+1) < f_i(k)$. Reversely, if $\theta_i(k) > f_i(k)$ then $\theta_i(k+1) > \theta_i(k)$ and $\theta_i(k+1) > f_i(k)$. Finally, if $\theta_i(k) = f_i(k)$ then $\theta_i(k+1) = f_i(k)$. □

Proposition 1. *Let $i \in \mathcal{V}$ and assume that Assumption 1 holds. If $\left(q_p(k)\right)_{k \geq 0}$ and $\left(n_i^+(k)\right)_{k \geq 0}$ are stationary sequences with limit q_p^* and n_i^{+*}, respectively. Then the sequence of opinion $\left(\theta_i(k)\right)_{k \geq 0}$ converges to*

$$\theta_i^* = \lim_{k \to \infty} \theta_i(k) = (1-\beta) \frac{2n_i^{+*} - n_i}{n_i} + \beta q_p^* \in \mathbb{Q} + \beta \mathbb{Q}$$

Proof. Let us note $f_i^* := (1-\beta)\left(2n_i^{+*} - n_i\right)/n_i + \beta q_p^* \in [-1, 1]$ and remarks that n_i^{+*} and $q_p^* \in \mathbb{N}$ since those are stationary sequences in $\mathbb{N}$. Then $f_i^* \in \mathbb{Q} + \beta\mathbb{Q}$. With Assumption 1, we have $\theta_i(0) \in (-1,1)$, so that Lemma 1 applies. Then by induction

$$\forall k \geq k^*, \quad \theta_i(k) \leq \theta_i(k+1) \leq f_i^* \leq 1$$

or

$$\forall k \geq k^*, \quad \theta_i(k) \geq \theta_i(k+1) \geq f_i^* \geq 1.$$

Then since $\left(\theta_i(k)\right)_{k \geq 0}$ is a bounded monotonous sequence, it converges. Let denote θ_i^* that limit. Now from above inequalities, if $\theta_i^* = 1$ then $f_i^* = 1$. Conversely, if $\theta_i^* = -1$, we have that $f_i^* = -1$. Finally, if $\theta_i^* \in (-1,1)$ then (8) rewrite as

$$\frac{\theta_i(k+1) - \theta_i(k)}{(1 - \theta_i(k)^2)} + \theta_i(k) = f_i^*.$$

Taking the limit of the previous cancel the first term of the left-hand side and we have that $\theta_i^* = f_i^*$. □

We will later see that opinions can either have an oscillatory chaotic behavior or they converge to a limit cycle. For a discrete-time system given by $x(k+1) = H(x(k))$, if there is a natural number $m > 1$ for which there exist m successive convergent sub-sequences $x_0(k) = x(mk)$, $x_1(k) = x(mk+1)$, ..., $x_{m-1}(k) = x(m(k+1)-1)$, then the overall sequence converges to a limit cycle of length m defined by the limits of the m sub-sequences.

Corollary 1. *Under Assumption 1, if $(q_p(k))_{k\geq 0}$ is a stationary sequence and $\left(n_i^+(k)\right)_{k\geq 0}$ converges to a limit cycle of length m, then θ_i also converges to a limit cycle of length m denoted $\bar{\theta}_i \subset \mathbb{Q} + \beta\mathbb{Q} + \cdots + \beta^m\mathbb{Q}$.*

B. Asymptotic behavior of the external/pollution dynamics

As pointed out in the previous subsection, the stationarity of $(q_p(k))_{k\geq 0}$ plays a major role in the asymptotic behavior of the opinions $\boldsymbol{\theta}$. Consequently, in this subsection we provide a sufficient condition ensuring that $(q_p(k))_{k\geq 0}$ is a stationary sequence. In this setting, the opinions behave as in the CODA model provided in [12] since the players are influenced by the external dynamics uniformly with respect to time after the sequences become stationary. We can rewrite the dynamics (2) by injecting (1)

$$p(k+1) = \gamma p(k) + n^+(k)e_{\max} + n^-(k)e_{\min} \tag{9}$$

Therefore, the pollution p reaches an equilibrium only if $(n^+(k))_{k\geq 0}$ (and implicitly $(n^-(k))_{k\geq 0}$) is stationary. Let us suppose that

$$\lim_{k\to\infty} n^+(k) = n^+, \quad \lim_{k\to\infty} n^-(k) = n^-$$

In this case the equilibrium p^* is given by

$$p^* = \frac{n^+ e_{\max} + n^- e_{\min}}{1-\gamma}, \tag{10}$$

which now only depends on the partition of actions of the individuals in the social network.

In order to guarantee that $(q_p(k))_{k\geq 0}$ is stationary one needs to ensure that $(\operatorname{sgn}(p(k)-\bar{p}))_{k\geq 0}$ is stationary. In other words, for k sufficiently large the value of $p(k)$ does not cross the threshold $\bar{p}$.

Lemma 2. *The sequence $(q_p(k))_{k\geq 0}$ is stationary for any graph $\mathcal{G}$ with N individuals if there exists k such that either $\left(p(k) \leq \frac{Ne_{max}}{1-\gamma} \wedge \frac{Ne_{max}}{1-\gamma} \leq \bar{p}\right)$ or $\left(p(k) \geq \frac{Ne_{min}}{1-\gamma} \wedge \frac{Ne_{max}}{1-\gamma} \geq \bar{p}\right)$.*

C. Preservation of action

In the following, we investigate under which condition we have that $q_i(k) = q_i(k+1)$. The following result is instrumental for our purposes.

Lemma 3. *Let $i \in \mathcal{V}$, then the following statements hold true:*

1) *if $f_i(k) \geq 0$ and $q_i(k) = 1$ then $q_i(k+1) = 1$,*
2) *if $f_i(k) \leq 0$ and $q_i(k) = -1$ then $q_i(k+1) = -1$.*

Proof. As proven in Lemma 1 one of (5), (6) or (7) holds true.
1) If (5) is verified one has $\theta_i(k+1) > \theta_i(k)$ meaning that $q_i(k) = 1$ implies $q_i(k+1) = 1$.
If (6) or (7) holds, then $\theta_i(k) \geq f_i(k)$. Consequently, $f_i(k) \geq 0$ ensures $\theta_i(k+1) \geq 0$ or equivalently $q_i(k+1) = 1$.
2) If (5) or (7) holds, then $\theta_i(k) \leq f_i(k)$. Therefore, if $f_i(k) < 0$ yields $\theta_i(k+1) < 0 \Leftrightarrow q_i(k+1) = -1$.
If (6) holds, one has that $\theta_i(k+1) < \theta_i(k)$ meaning that $q_i(k) = -1$ implies $q_i(k+1) = -1$.
□

When $\beta < 1/(1+n_i)$ Lemma 3 can be refined as follows.

Lemma 4. *Let $i \in \mathcal{V}$ and assume that $\beta < 1/(1+n_i)$. The following statements hold:*
– if $n_i^+(k) > n_i^-(k)$ and $q_i(k) = 1$ then, $q_i(k+1) = 1$,
– if $n_i^-(k) < n_i^+(k)$ and $q_i(k) = -1$ then, $q_i(k+1) = -1$.

Proof. Notice that $\beta < 1/(1+n_i)$ implies $\dfrac{\beta}{1-\beta} < \dfrac{1}{n_i}$. Recalling that $|q_p(k)| = 1$ one obtains that $\left|\dfrac{n_i q_p(k)\beta}{1-\beta}\right| < 1$. Notice also that

$$\begin{aligned} f_i(k) &= (1-\beta)\frac{n_i^+(k) - n_i^-(k)}{n_i} + \beta q_p(k) \\ &= \frac{1-\beta}{n_i}\left(n_i^+(k) - n_i^-(k) + \frac{n_i q_p(k)\beta}{1-\beta}\right) \end{aligned}$$

Therefore

$$\operatorname{sgn}(f_i(k)) = \operatorname{sgn}\left(n_i^+(k) - n_i^-(k) + \frac{n_i q_p(k)\beta}{1-\beta}\right),$$

and taking into account that $n_i^+(k), n_i^-(k) \in \mathbb{N}$ the desired result yields from Lemma 3. □

Throughout the paper, we denote by $|A|$ the cardinality of a set A. We provide a definition for some cluster in the graph such that the opinion will not change through time

Definition 1. *We say that a subset of agents $A \subset \mathcal{V}$ is a weakly robust polarized cluster if the following hold:*

- $\forall i, j \in A,\ q_i(0) = q_j(0)$,
- $\forall i \in A,\ |\mathcal{N}_i \cap A| \geq |\mathcal{N}_i \setminus A| - \beta/(1-\beta)|\mathcal{N}_i|$.

Proposition 2. *If A is a weakly robust polarized cluster and $q_p(k) = q_i(0)$ for all $k \in \mathbb{N}$ and $i \in A$, then*

$$q_i(k) = q_i(0),\ \forall i \in A, \forall k \in \mathbb{N}.$$

Proof. The proof will be done by induction. Let us suppose that $\forall i \in A$ one has $q_i(0) = 1$. Assume that for a given $k' \in \mathbb{N}$ one has $q_i(k') = 1,\ \forall i \in A$. Since the interaction graph is fixed and A is a weakly robust polarized cluster the following holds true $|\mathcal{N}_i \cap A| \geq |\mathcal{N}_i \setminus A| - \beta/(1-\beta)|\mathcal{N}_i|$. Noticing that $\mathcal{N}_i \cap A \subseteq \mathcal{N}_i^+(k')$ one obtains that $n_i^+(k') \geq n_i^-(k') - n_i\beta/(1-\beta)$ which is equivalent to $f(k') \geq 0$. Applying Lemma 3 one gets $q_i(k'+1) = 1$ and the induction is complete. Similar reasoning applies when $q_i(0) = -1,\ \forall i \in A$. □

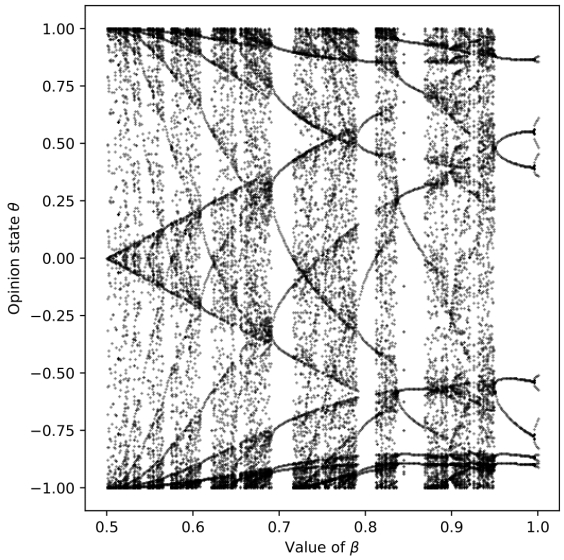


Fig. 1: Bifurcation diagram of the opinion for $0.5 < \beta < 1$. $N = 20$, $\theta(0) = 0.4$, $p(0) = 100$, $\bar{p} = 15$, $e_{\min} = 0$, $e_{\max} = 1$ and $\gamma = 0.5$,

The preservation of action in a weakly polarized cluster is subject to a constant value of q_p over the time. In order to get rid of this constraint, we introduce the following concept.

Definition 2. *We say that a subset of agents $A \subset \mathcal{V}$ is a strongly robust polarized cluster if the following hold:*

- $\forall i, j \in A,\ q_i(0) = q_j(0)$,
- $\forall i \in A,\ |\mathcal{N}_i \cap A| \geq |\mathcal{N}_i \setminus A| + \beta/(1-\beta)|\mathcal{N}_i|$.

Proposition 3. *If A is a robust polarized cluster then $\forall i \in A$, $\forall k \in \mathbb{N},\ q_i(k) = q_i(0)$.*

Proof. Apply proof of Proposition 2 while considering $|\mathcal{N}_i \cap A| \geq |\mathcal{N}_i \setminus A| + \beta/(1-\beta)|\mathcal{N}_i|$. □

It is worth noting that for $\beta > 1/2$ we cannot have robust polarized cluster since $\beta/(1-\beta) > 1$ and $|\mathcal{N}_i \cap A| \leq \beta/(1-\beta)|\mathcal{N}_i|$ for any $A \subset \mathcal{V}$. This fact will have importance in the following.

IV. Analysis of the synchronized behavior

In the case where $\beta > 1/2$ the state of the system does not converge towards a steady state. Instead, one has oscillations that may be either chaotic or converging to a limit cycle. This is illustrated in Fig. 1 where we can see that for β close to one we get a limit cycle while for $\beta > 1/2$, in a wide range, one has a chaotic behavior. For the sake of simplicity, we assume in the following that all the opinions are synchronized.

Definition 3. *We say that an opinion state $\boldsymbol{\theta} = (\theta_1 \cdots \theta_N)^\top$ is Fully Synchronized (FS) if $\forall i, j \in \mathcal{V},\ \theta_i = \theta_j$.*

When the opinion state is FS, the opinion of an agent $i \in \mathcal{V}$ is equal to the opinion of any other agent in $\mathcal{V}$. Therefore, in the remainder of this section, we will denote $\theta(k)$ and $q(k)$ the common opinion and action of all the agents at time k. In other words, we omit the agent index when referring to its opinion or action.

Proposition 4. *The FS property is forward invariant over time, i.e. if $\boldsymbol{\theta}(k)$ is FS at time $k \in \mathbb{N}$, then $\boldsymbol{\theta}(k+1)$ is FS.*

Remark 1. *If $\boldsymbol{\theta}(k)$ is FS then $f(k) = (1-\beta)q(k) + \beta q_p(k)$.*

Indeed, under the assumption of FS, one has that for all $i \in \mathcal{V}$, $(n_i^+(k) - n_i^-(k))/n_i = q_i(k) = q(k)$.

In the FS regime, the action space reduces to

$$\begin{aligned}\mathcal{S} &= \{(q, q_p) \mid q \in \{-1, 1\}, q_p \in \{-1, 1\}\} \\ &= \{(-1,-1), (-1,1), (1,-1), (1,1)\}.\end{aligned}$$

Notice that each point in $\mathcal{S}$ corresponds to a partition of the state space in four sets. For instance $(1,1)$ corresponds to $\{\theta_i(k) \geq 0, \forall kp \leq \bar{p}\}$ In order to prove the oscillatory behavior of the system (2)-(3) we show that in general $\mathcal{S}$ does not contain equilibrium points. This means that the trajectory of the system cannot remain in a certain partition which means it cannot converge towards a steady state.

Proposition 5. *Assume that $\beta > 1/2$ and the opinion state is FS at time $k \in \mathbb{N}$. Then the points $\{(1,-1), (1,-1)\}$ are not equilibrium in the action space.*

Proof. We proceed by contradiction. The reasoning is similar for each of the two points so we will focus on the first one. Let us assume that $(1,-1)$ is an equilibrium point. This means that if there exists $k^* \in \mathbb{N}$ such that $q(k^*) = 1,\ q_p(k^*) = -1$ then for any $k > k^*$ we have $q(k) = 1$ and $q_p(k) = -1$. We notice that in this case, one has $f(k) = 1 - 2\beta,\ \forall k > k^*$. Then the dynamics (8) becomes

$$\theta(k+1) = \theta(k) + \left(1 - \theta(k)^2\right)\left(1 - 2\beta - \theta(k)\right),\ \forall k > k^*.$$

Recalling that $q(k) = 1,\ \forall k > k^*$ one deduces that $\theta(k) \geq 0,\ \forall k > k^*$. Consequently, $\forall k > k^*$ one obtains

$$\theta(k+1) - \theta(k) = \left(1 - \theta(k)^2\right)\left(1 - 2\beta - \theta(k)\right) \leq 1 - 2\beta.$$

Iterating the inequality above for consecutive values of k it results that for any $n \in \mathbb{N}$ the following holds:

$$\theta(k+n) \leq n(1-2\beta) + \theta(k) \leq n(1-2\beta) + 1. \quad (11)$$

Let us recall that $1 - 2\beta < 0$. Therefore, if we consider in (11) a sufficiently large n (i.e. $n > 1/(2\beta - 1)$) we get $\theta(k+n) < 0$ yielding$q(k+n) = -1$ which contradicts the assumption that $(1,-1)$ is an equilibrium in the action space. □

Proposition 6. *Assume $\beta > 1/2$ and $\boldsymbol{\theta}$ being FS. The following statements hold true.*

- *$(1,1)$ is an equilibrium in the action space if and only if $p_{\max} \leq \bar{p}$ where $p_{\max} = Ne_{max}/(1-\gamma)$. In this case $(\boldsymbol{\theta}, p) = (\mathbf{1}_n, p_{\max})$ is a stable equilibrium of the coupled dynamics* (2)-(3).
- *$(-1,-1)$ is an equilibrium in the action space if and only if $p_{\min} \geq \bar{p}$ where $p_{\min} = Ne_{min}/(1-\gamma)$. In this*

case $(\boldsymbol{\theta}, p) = (-\mathbf{1}_n, p_{\min})$ *is a stable equilibrium of the coupled dynamics* (2)-(3)*.*

Proof. The reasoning is symmetric and it is sufficient to focus only on the first statement.
"⇒" We assume that $(1,1)$ is an equilibrium and show that $p_{\max} < \bar{p}$.
Suppose that there exists $k^* \in \mathbb{N}$ such that $q(k^*) = 1,\ q_p(k^*) = 1$. Then, for all $k > k^*$, we have $q(k) = 1$ and $q_p(k) = 1$. Therefore, for all $k > k^*$ one has $p(k) \leq \bar{p}$ and the dynamics (2) becomes

$$p(k+1) = \gamma p(k) + N e_{\max}. \tag{12}$$

Since $\gamma \in (0,1)$ the dynamics above is asymptotically stable and $p(k)$ converges to the equilibrium defined in (10) with $n^+ = N,\ n^- = 0$ i.e. $p^* = p_{\max} = N e_{\max}/(1-\gamma)$. We conclude that $p_{\max} \leq \bar{p}$.
"⇐" We assume that $p_{\max} \leq \bar{p}$ and show that $(1,1)$ is an equilibrium.
One proves that if k is such that $q(k) = 1,\ q_p(k) = 1$ then $q(k+1) = 1,\ q_p(k+1) = 1$. Notice first that in the case under consideration $f(k) = 1$. Therefore, dynamics (3) becomes

$$\theta(k+1) = \theta(k) + (1 - \theta(k)^2)(1 - \theta(k)) \geq \theta(k). \tag{13}$$

Since $q(k) = 1$ one deduces from (13) that $q(k+1) = 1$ as well.
On the other hand $q_p(k) = 1$ implies $p(k) < \bar{p}$ and $p_{\max} \leq \bar{p}$ is equivalent to $N e_{\max} \leq (1-\gamma)\bar{p}$. Therefore, (12) becomes

$$p(k+1) < \gamma\bar{p} + (1-\gamma)\bar{p} = \bar{p}$$

yielding $q(k+1) = 1$. By recursive reasoning one gets that $(1,1)$ is an equilibrium point.
We already noticed that in the case under study $p_{\max}$ is an asymptotically stable point for (2) that takes the particular form (12). On the other hand the dynamics (3) simplifies as (13) whose stable equilibrium is $\mathbf{1}_n$. □

In the general case when $p_{min} < \bar{p} < p_{\max}$ neither $(1,1)$ nor $(-1,-1)$ is an equilibrium in the action space. Since no equilibrium exists in this case, the trajectory of (2)-(3) will switch an infinite number of times between the four sets of the partition defined by the action space $\mathcal{S}$.

V. Numerical results

First, we will illustrate the same kind of result as in [12] over a square lattice. Finally, we will present the different behaviors we can observe for a complete graph with FS property.

A. Square lattice

Our study visualizes results with interactions based on a square lattice topology. In Figure 2, we note the persistence of resilient clusters even after numerous iterations. The evolution of opinions and the corresponding state through iterations is depicted in Figure 3. We can observe that opinions and state converge fast to a limit cycle. On Figure 2, we see that the opinions are polarized on the graph. There

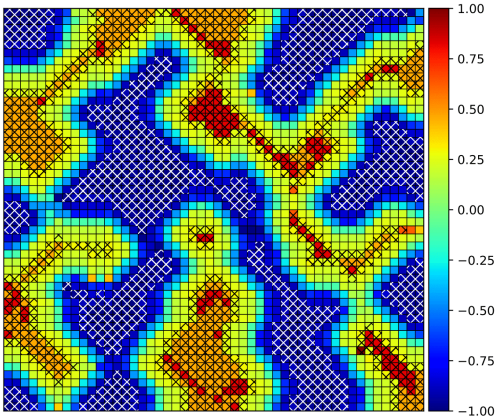


Fig. 2: Visualization of Opinion Dynamics on a 50 × 50 Square Lattice for $\beta = 0.45$: Initial opinions are randomly distributed as i.i.d. uniform variables between -1 and 1, with the resultant opinions after 100 iterations represented by each colored square cell. Agents engage in communication with their adjacent cells (above, below, left, and right). The cells marked with crosses indicate the presence of strongly robust polarized clusters; black crosses correspond to action -1, and white crosses denote action 1.

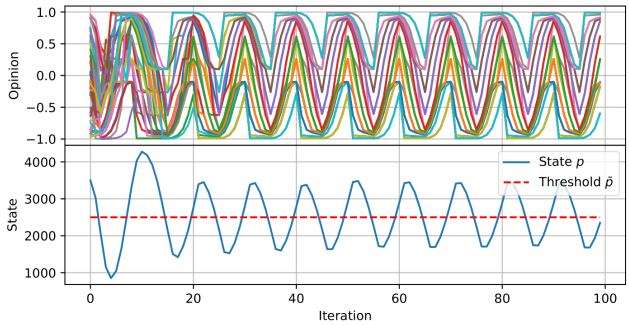


Fig. 3: Depiction of Dynamical Evolution in the 50 × 50 Square Lattice from Figure 2: The upper panel showcases the trajectory of each agent's opinion over iterations, while the lower panel illustrates the corresponding state evolution. The red dashed line marks the state threshold.

are many robust clusters for both action 1 and -1. Each of them is separated by a frontier that seem to have the same length between the clusters of opposite action and them. We can see on Figure 3 that there is no agent with a constant opinion, the opinions are on a limit cycle. The ones that stay with a constant action form robust clusters, as illustrated in Figure 2.

B. Complete graph with FS

In Figures 4, we identify three unique behaviors displayed by the dynamical systems: stable equilibrium (when $\beta < 0.5$), chaotic patterns, and a collection of limit cycles. These simulations were conducted on a complete graph comprising 20 nodes, all of which possess the FS property at initial time. The starting opinion is set at $\theta(0) = 0.4$, and the initial

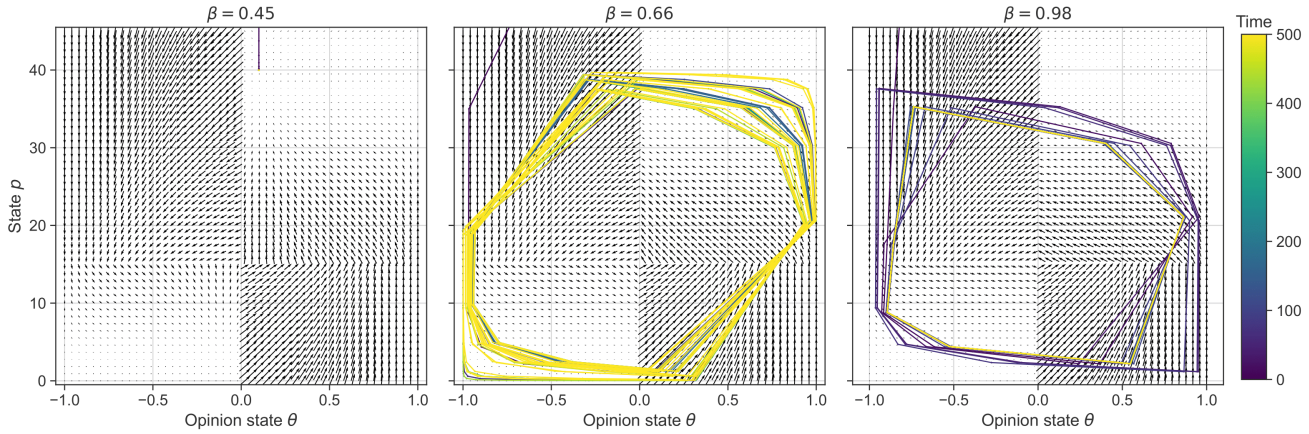


Fig. 4: Trajectory dynamics. Left: two possible equilibria. Center: no equilibrium or limit cycle. Right: limit cycle.

state is taken at $p(0) = 100$, with a threshold of $\bar{p} = 15$. For the emission dynamics, the range is between $e_{\min} = 0$ and $e_{\max} = 1$, and the decay rate is defined by $\gamma = 0.5$. We represent the discrete trajectory linking each consecutive couple $(\theta(k), p(k))$ by a straight line. The color of the line represents the time when this shift occurs. Moreover, we present the subsequent vector field of the dynamics illustrated by the quivers. The speed of the dynamics is given by the length of the arrows.

A quick observation reveals that, based on the given parameters, the dynamics of the opinion and state undergo significant variations in response to the value of β. For instances when $\beta < 0.5$, both the opinion and state quickly stabilize at an equilibrium. However, as depicted in Figure 1, given the same conditions, when $\beta > 0.5$ we discern two potential behaviors. The first sees β positioned outside all limit cycle intervals, resulting in chaotic behavior as illustrated in the center of Figure 4. The second showcases a limit cycle, as depicted in the right of Figure 4. Specifically, in the Chaos case of Figure 4, the sequence $((\theta(k), p(k)))_{k\geq 0}$ never reiterates, thus forming its unique pattern. Conversely, when β falls within a limit cycle interval, the sequence $((\theta(k), p(k)))_{k\geq 0}$ converge towards a limit cycle.

VI. Conclusion

In this paper, we have introduced and analyzed a CODA model coupled with an external dynamic state. Specifically, we consider that the external dynamics represent a very simple pollution model in which the emission level depends on the actions of the individuals in the social network. Conversely, the opinions are both influenced by the actions of the neighbors and the pollution level (above or below a given threshold). We have shown that different behaviors are possible, ranging from convergence to a steady state to chaotic behavior of the coupled dynamics. Numerical examples illustrate our theoretical results.

References

[1] E. Ising, "Contribution to the theory of ferromagnetism," Ph.D. dissertation, University of Hamburg, 1924.

[2] P. Clifford and A. Sudbury, "A model for spatial conflict," *Biometrika*, vol. 60, no. 3, pp. 581–588, 1973.

[3] K. Sznajd-Weron and J. Sznjad, "Opinion evolution in closed community," *International Journal of Modern Physics C*, vol. 11, no. 6, pp. 1157–1165, 2000.

[4] M. A. V.S. Varma, I.C. Morărescu, "Analysis of opinion dynamics under binary exogenous and endogenous signals," *Nonlinear Analysis: Hybrid Systems*, vol. 38, p. 100910, 2020.

[5] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Advances in Complex Systems*, vol. 3, pp. 87–98, 2000.

[6] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 3, 2002.

[7] N. E. Friedkin and E. C. Johnsen., "Social influence and opinions." *Journal of Mathematical Sociology.*, vol. 15, pp. 193–206, 1990.

[8] I.-C. Morărescu and A. Girard, "Opinion dynamics with decaying confidence: Application to community detection in graphs," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1862 – 1873, 2011.

[9] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[10] C. Altafini, "Consensus problems on networks with antagonistic interactions," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 935–946, 2013.

[11] A. Martins, "Continuous opinions and discrete actions in opinion dynamics problems," *International Journal of Modern Physics C*, vol. 19, no. 4, pp. 617–625, 2008.

[12] N. R. Chowdhury, I.-C. Morărescu, S. Martin, and S. Srikant, "Continuous opinions and discrete actions in social networks: a multi-agent system approach," in *Proceedings 55th IEEE Conference on Decision and Control*, 2016, pp. 1739–1744.

[13] F. Ceragioli and P. Frasca, "Continuous-time consensus dynamics with quantized all-to-all communication," in *Proceedings of European Control Conference*, 2015.

[14] R. A. Memon, D. Y. C. Leung, and C.-H. Liu, "Effects of building aspect ratio and wind speed on air temperatures in urban-like street canyons," *Building and Environment*, vol. 45, no. 1, pp. 176–188, Jan. 2010.

[15] R. Berkowicz, "OSPM - A Parameterised Street Pollution Model," *Environmental Monitoring and Assessment*, vol. 65, no. 1, pp. 323–331, Nov. 2000.

[16] J. Baker, H. L. Walker, and X. Cai, "A study of the dispersion and transport of reactive pollutants in and above street canyons—a large eddy simulation," *Atmospheric Environment*, vol. 38, no. 39, pp. 6883–6892, Dec. 2004.

[17] A. Garg, A. Kumar, and N. C. Gupta, "Comprehensive study on impact assessment of lockdown on overall ambient air quality amid COVID-19 in Delhi and its NCR, India," *Journal of Hazardous Materials Letters*, vol. 2, p. 100010, Nov. 2021.

Analysis of an opinion dynamics model coupled with an external environmental dynamics

Anthony Couthures[a], Vineeth Satheeskumar Varma[a,b], Samson Lasaulce[a,c], Irinel - Constantin Morărescu[a,b]

[a] *Université de Lorraine, CNRS, CRAN, Nancy, F-54000, France*
[b] *Automation Department, Technical University of Cluj-Napoca, Cluj-Napoca, Romania*
[c] *Khalifa University, Abu Dhabi, UAE*

Abstract

We consider a set of individuals, referred to as agents, whose opinions evolve according to nonlinear dynamics. Their opinions impact their behavior or actions, which in turn affect their local environment (for example, via pollution, contamination of a virus, etc.). Each agent can also perceive or observe a signal about the environment, and is influenced by this external signal. This yields a coupled dynamics (opinion and external signal), which behaves in a similar manner to the prey-predator models. One of the main features of our study is that the information provided by the external signal has a significant impact on the opinion dynamics. When the coupling is strong, the external signal may induce either chaotic behavior or convergence towards a limit cycle. When the coupling with the external signal is weak, the classical behavior characterized by local agreements in polarized clusters is observed. In both cases, conditions under which clusters of individuals don't change their actions are provided. Numerical examples are provided to illustrate the derived analytical results.

Keywords:
Complex networks, Opinion dynamics, Bifurcations, Chaotic Systems, Limit cycles

1. Introduction

Opinion dynamics (OD) over social networks is a topic that has received increasing attention during the last few decades. One effective way to model OD under social interactions is through multi-agent systems formalism. Several OD models exist, and they can be split in different ways. The first one assumes that opinions evolve in a discrete set of values [1, 2] while others consider that opinions lie in a continuous set of values [3, 4, 5, 6]. While some models naturally lead to consensus [5, 7], others yield network clustering [3, 4, 6, 8]. A common feature of all the models enumerated above is that each individual has access to the exact opinion values of its neighbors. This assumption is not realistic, and to overcome it, a mix of continuous opinion with discrete actions (CODA) was proposed in [9]. This model reflects the fact that even if we often face binary choices or actions that are visible to our neighbors, our opinion evolves in a continuous space of values that are not accessible. A consensus-like dynamics reproducing this behavior has been proposed and analyzed in [10] where the preservation and propagation of actions are also characterized through the notion of robust polarized clusters. While the model in [10] led to a clustering of the network, a similar idea was employed in [11] to study the emergence of consensus under quantized all-to-all communication.

In this paper, we analyze the behavior of a slightly generalized CODA model coupled with the dynamics of an external environmental state, which represents, in this work, the atmospheric pollution levels. It is noteworthy that our results can be adapted to other scenarios in which the external signal describes virus contamination or marketing instances. The usage of the CODA model for opinion dynamics is justified by two important social features: 1) individuals may have access to an inexact/quantized perception of the opinion of their neighbors, 2) the extremists are more stubborn and their opinion is more difficult to change. Many models have been developed to characterize the atmospheric pollution dynamics[12]. One of the popular modeling approaches is based on the half-life decomposition of polluting compounds (see, for instance, [13, 14]). Generally, the time constants of the processes depend on the chemical compound considered [14, 15] and the dynamics may be cumbersome. Since our focus is not on the precise modeling of this process, we use a simple linear pollution model to describe the evolution of the environmental air pollution. In this model, the state of the environment depends on the states (capturing the environmental behavior) of the individuals, which in turn are influenced both by the states of their neighbors and the external signal (pollution level). The coupling of the two dynamics leads to a complex asymptotic behavior that can be summarized as follows: When the coupling between the dynamics is weak, one recovers the asymptotic behavior similar to the CODA model in [10] or to [16]. A strong coupling between the two dynamics hampers the convergence towards a steady state and yields either chaotic oscillation or convergence towards a limit cycle. It is noteworthy that even in the simplified case, when all the agents have the same initial opinion, the strong coupling with the external dynamics hampers the convergence toward a steady state and may lead to chaotic oscillations. As shown in the sequel, our model behaves as a prey-predator model where the predators thrive when the prey thrive, but this leads to the decline of the prey, which in turn yields the decline

of the predators [17, 18].

We point out that the idea of coupling environment dynamics and population behavior is not new, it has been mainly studied in the framework of evolutionary games [19, 20, 21]. While these studies also emphasize similar oscillatory asymptotic behavior, they employ different tools and the insights are different. For example, in [19, 20, 21] the propagation of the information is homogeneous (the entire population gets the same information) and the oscillations are induced by the time-varying proportion of the population adopting one of the two possible strategies. In our case, the information propagates through a graph of individuals with different stubbornness and the oscillations are induced by the pollution level that triggers different social behaviors.

The main contributions of this paper are: i) the introduction of a mathematical model capturing the coupling between a nonlinear opinion dynamics model and an external signal; ii) the analysis of the asymptotic behavior of the aforementioned model. Those contributions follow from the work initiated in [22]. Note that both the opinion dynamics and the pollution dynamics models are inspired by existing works in the literature.

The paper is structured as follows: Section 2 introduces the main notions and concepts necessary to describe the mathematical model that is analyzed in the sequel. Characteristics of equilibrium and asymptotic behavior are analyzed in Section 3. Section 5 numerically highlights various setups leading to different asymptotic behaviors. The manuscript ends with Section 6 which presents concluding remarks about the work.

In the following, $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ denote the set of real numbers and non-negative real numbers, respectively. We will use the notation for the identity matrix $\mathbf{I} \in \mathbb{R}^{N\times N}$ when there is no ambiguity. For the specific case of dimension $n \in \mathbb{N}$, we denote the identity matrix as $\mathbf{I}_n \in \mathbb{R}^{n\times n}$. The scalar identity function is denoted as $\mathrm{Id} : \mathbb{R} \to \mathbb{R}$, where $\mathrm{Id}(x) = x$. Moreover, we denote by $\mathbf{1}$ the all-ones vector and by $\mathbf{0}$ the null vector in $\mathbb{R}^N$ when there is no ambiguity, and for the n-dimensional cases, $\mathbf{1}_n$ and $\mathbf{0}_n$ respectively. In the sequel, for $i \in \mathcal{V}$, e_i denotes the i-th element of the canonical basis of $\mathbb{R}^N$. For a function $f : X \to X$, we denote by $\mathrm{Fix}(f) \subset X$ the set of fixed points of f, i.e., $\mathrm{Fix}(f) := \{x \in X \mid f(x) = x\}$.

2. Problem formulation and preliminaries

We consider the classical multi-agent framework in which N individuals/agents belonging to the set $\mathcal{V} = \{1, \dots, N\}$ and interacting according to a fixed *directed graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ which does not contains loop. We denote by $\boldsymbol{A} \in {\mathbb{R}_+}^{N\times N}$ its adjacency matrix, such that $A_{ij} = 1$ if $(i, j) \in \mathcal{E}$ and 0 otherwise, by $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ its degree matrix and by $\boldsymbol{L} \in \mathbb{R}^{N\times N}$ where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ the associated Laplacian. The neighborhood of the agent i is denoted by $\mathcal{N}_i$ and represents the set of agents that influence i according to the graph $\mathcal{G}$ (i.e $j \in \mathcal{N}_i \Leftrightarrow (j, i) \in \mathcal{E}$). We denote by n_i the cardinality of $\mathcal{N}_i$. Let us also recall that a path in $\mathcal{G}$ is a finite sequence of edges $(i_1, i_2), (i_2, i_3), \dots, (i_p, i_{p+1})$ such that $(i_k, i_{k+1}) \in \mathcal{E}$ for all $k \in \{1, \dots, p\}$. Two vertices $i, j \in \mathcal{V}$ are *connected* in $\mathcal{G}$ if there exists a path in $\mathcal{G}$ joining i and j (i.e. $i_1 = i$ and $j_p = j$). Finally, a directed graph $\mathcal{G}$ is strongly connected if any two nodes in $\mathcal{V}$ are connected.

At each time instant $k \in \mathbb{N}$ we assign to each agent $i \in \mathcal{V}$ an opinion $\theta_i(k) \in \Theta := [-1, 1]$ that evolves according to the discrete-time protocol defined further in (2). We also introduce $\boldsymbol{\theta}(k) = (\theta_i(k))_{i\in\mathcal{V}} \in \boldsymbol{\Theta} := [-1, 1]^N$ the vector collecting the opinions of all the agents at time k. Note that $\theta_i(k)$ is an abstraction which is closer to -1 when the individual i is more convinced to have a pro-environmental behavior and closer to 1 when the individual does not care about the environment and behaves accordingly. We assume that the only information from the neighbors is the behavior/action captured by the monotonically increasing function $s_\theta : \Theta \to \Theta$ as a proxy of θ_i and verifying $s_\theta(-1) = -1$ and $s_\theta(1) = 1$. Let us denote $\boldsymbol{s}_\theta(\boldsymbol{\theta}) = (s_\theta(\theta_i))_{i\in\mathcal{V}}$.

Remark 1. *Note that s_θ can be the identity function meaning that the agents can access the actual opinions of their neighbors. The other extreme case is when s_θ is when the* sign *function in which the agents can only detect if the neighbors are above or below the average of the eco-responsibility. Moreover, the intermediate situation is also possible with s_θ being a smooth sigmoid function like $s_\theta(\theta) = 2\arctan(m\theta/(1-\theta^2))/\pi$ with $m \in \mathbb{N}$ a parameter allowing to change the steepness of the curve.*

We assume that the opinion of agent i at time k induces a behavior generating certain emissions, which are captured by the increasing function $E : \Theta \to [e_{\min}, e_{\max}]$ where $0 < e_{\min} < e_{\max}$ are the minimum and maximum emissions, respectively. For simplicity, we note the emission of agent i at time k with $E(\theta_i(k))$. Also, we denote by $\boldsymbol{E}(\boldsymbol{\theta}) = (E(\theta_i))_{i\in\mathcal{V}}$ the vector of emission of all the agents.

The state of the environment is characterized by another abstraction, $p \in \mathbb{R}_{\geq 0}$, that captures the pollution state under the emission of the overall network of individuals. To keep things simple, we assume that the pollution evolves according to the following discrete-time dynamics:

$$\begin{aligned} p(k+1) &= (1-\gamma)\, p(k) + \sum_{i=1}^{N} E(\theta_i(k)) \\ &= (1-\gamma)\, p(k) + \mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}(k)), \end{aligned} \tag{1}$$

where $\gamma \in (0, 1)$ is an autonomous decay rate. This dynamics is taken as a simplification of equation (2) of [14].

We also assume that individuals cannot observe p accurately, instead, they can perceive an output of the environmental dynamics defined by the function $s_p : \mathbb{R} \to \Theta$. The output s_p of the air pollution is relative to a deviation from a threshold denoted $\bar{p} \in \mathbb{R}_{\geq 0}$. Basically, the population will perceive a polluted environment if $p > \bar{p}$ and will have the feeling that air is healthy otherwise. The function s_p is increasing with respect to the difference $\bar{p} - p$ and such that $s_p(q) = 0$ if and only if $q = 0$.

We also state the following standing assumption:

Standing Assumption 1. *For $\gamma \in (0, 1)$ such as in* (1)*, we have $Ne_{min}/\gamma < \bar{p} < Ne_{max}/\gamma$.*

This assumption allows us to exclude the case when the external signal has no significant influence on the opinion dynamics. When this assumption is not satisfied the analysis reduces to the OD dynamics, as even when all the agents are polluting (or not), the air pollution signal remains below (or above) the threshold. Further details on this case can be found in [22].

We are now ready to describe the OD model that we consider in this work. This dynamics adapts the CODA model in [10] and [16] to include the external dynamics of $p(k)$:

$$\theta_i(k+1) = \theta_i(k) + \left(1-\theta_i(k)^2\right)\Big[\beta\left(s_p(\bar{p}-p(k)) - \theta_i(k)\right) + (1-\beta)\,\frac{1}{n_i}\sum_{j\in\mathcal{N}_i}\left(s_\theta(\theta_j(k)) - \theta_i(k)\right)\Big], \tag{2}$$

where $0<\beta<1$ encapsulates the trade-off between the external signal and the neighbors' behavior on the opinion evolution.

Noting the state as $\boldsymbol{x}(k) = (\boldsymbol{\theta}(k), p(k))$, we can rewrite (1) and (2) in the collective form as:

$$\boldsymbol{x}(k+1) = \mathbf{F}(\boldsymbol{x}(k)) = \begin{pmatrix}\boldsymbol{u}(\boldsymbol{\theta}(k), p(k))\\ v(\boldsymbol{\theta}(k), p(k))\end{pmatrix}, \tag{3}$$

where the functions are,

$$\boldsymbol{u}(\boldsymbol{\theta}, p) = \boldsymbol{\theta} + \left(\mathbf{I} - \operatorname{diag}(\boldsymbol{\theta})^2\right)(\boldsymbol{f}(\boldsymbol{\theta}, p) - \boldsymbol{\theta}),$$
$$v(\boldsymbol{\theta}, p) = (1-\gamma)\,p + \mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}),$$

and

$$\boldsymbol{f}(\boldsymbol{\theta}, p) = \beta s_p(\bar{p}-p)\mathbf{1} + (1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}_\theta(\boldsymbol{\theta}).$$

Having established the foundational dynamics model, we now turn to a detailed analysis of its behavior.

3. Analysis of the general case

This section delves into an analysis of the model's dynamics, beginning with the characterization of equilibria in the general case and incorporating results that exploit the graph structure. Subsequently, we examine the stability of certain equilibrium points located at the extremes of the state space. Additionally, we explore scenarios involving fully synchronized opinions and study the system's equilibria under piece-wise linear signals.

But first, the following general invariance result can be stated.

Lemma 1. *The compact set $\mathcal{X} = \boldsymbol{\Theta}\times[Ne_{min}/\gamma, Ne_{max}/\gamma]$ is forward invariant under the dynamics* (3)*. i.e. $\forall \boldsymbol{x}\in\mathcal{X}$, $\boldsymbol{F}(\boldsymbol{x})\in\mathcal{X}$.*

Proof. Note that $\boldsymbol{f}(\boldsymbol{\theta}, p)$ is the convex combination of two vectors in $\boldsymbol{\Theta}$. Thus, for all $(\boldsymbol{\theta}, p)\in\boldsymbol{\Theta}\times\mathbb{R}_{\geq 0}$ and for any $i\in\mathcal{V}$ one obtains,

$$-1 \leq e_i^\top \boldsymbol{f}(\boldsymbol{\theta}, p) \leq 1 \Leftrightarrow -1-\theta_i \leq e_i^\top \boldsymbol{f}(\boldsymbol{\theta}, p) - \theta_i \leq 1-\theta_i,$$

which implies:

$$-1 \leq e_i^\top \boldsymbol{u}(\boldsymbol{\theta}, p) \leq 1 \Leftrightarrow \boldsymbol{u}(\boldsymbol{\theta}, p)\in\boldsymbol{\Theta}.$$

Furthermore, for any $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, the inequality $Ne_{\min} < \boldsymbol{E}(\boldsymbol{\theta}) < Ne_{\max}$ holds. Then, for any $(\boldsymbol{\theta}, p)\in\mathcal{X}$, one has,

$$\begin{aligned}&\frac{Ne_{\min}}{\gamma} < p < \frac{Ne_{\max}}{\gamma}\\ \Leftrightarrow\ &(1-\gamma)\frac{Ne_{\min}}{\gamma} + Ne_{\min} < v(\boldsymbol{\theta}, p) < (1-\gamma)\frac{Ne_{\max}}{\gamma} + Ne_{\max}\\ \Leftrightarrow\ &\frac{Ne_{\min}}{\gamma} < v(\boldsymbol{\theta}, p) < \frac{Ne_{\max}}{\gamma}.\end{aligned}$$

Therefore, for any $\boldsymbol{x}\in\mathcal{X}$, it follows that $\boldsymbol{F}(\boldsymbol{x})\in\mathcal{X}$. □

This Lemma ensures the forward invariance of the compact set $\mathcal{X}$ under the dynamics (3), meaning that for every trajectory $(x(k))_{k\in\mathbb{N}}$ such that there exists a $k\in\mathbb{N}$ with $x(k)\in\mathcal{X}$, then for all $l\in\mathbb{N}$, it follows that $x(k+l)\in\mathcal{X}$. Additionally, it can be observed that $\mathcal{X}$ is attractive with respect to the environment dimension, meaning that for any $x(0)\in\boldsymbol{\Theta}\times\mathbb{R}_{\geq 0}$, there exists a time $k\in\mathbb{N}$ such that $\boldsymbol{x}(k)\in\mathcal{X}$.

3.1. Characterization of equilibria

We start with the characterization of the equilibria of (3) i.e., the points $\boldsymbol{x}^*$ such that $\boldsymbol{x}^* = \mathbf{F}(\boldsymbol{x}^*)$ or $\boldsymbol{x}^*\in\operatorname{Fix}(\boldsymbol{F})$. Before introducing the following result let us define the instrumental function $\boldsymbol{g}:\boldsymbol{\Theta}\to\boldsymbol{\Theta}$ such that

$$\boldsymbol{g}(\boldsymbol{\theta}) = \beta s_p\left(\bar{p} - \frac{\mathbf{1}^\top\boldsymbol{E}(\boldsymbol{\theta})}{\gamma}\right)\mathbf{1} + (1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}_\theta(\boldsymbol{\theta}). \tag{4}$$

Proposition 1. *Let $\boldsymbol{x}^* = (\boldsymbol{\theta}^*, p^*)$ be an equilibrium point of the dynamics* (3)*. Then $\boldsymbol{x}^* = (\boldsymbol{\theta}^*, \mathbf{1}^\top\boldsymbol{E}(\boldsymbol{\theta}^*)/\gamma)$ such that $\boldsymbol{\theta}^*\in\mathcal{F}$ with*

$$\mathcal{F} = \left\{(\theta_i)_{i\in\mathcal{V}}\in\boldsymbol{\Theta} \mid \forall i\in\mathcal{V},\ \theta_i\in\{-1,1\}\ \text{or}\ \theta_i = e_i^\top\boldsymbol{g}(\boldsymbol{\theta})\right\}. \tag{5}$$

Proof. Assuming that $\boldsymbol{x}^*$ is an equilibrium of (3) it is straightforward to see from (1) that

$$p^* = v(\boldsymbol{\theta}^*, p^*) \Leftrightarrow p^* = (1-\gamma)\,p^* + \mathbf{1}^\top\boldsymbol{E}(\boldsymbol{\theta}^*) \Leftrightarrow p^* = \frac{\mathbf{1}^\top\boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}.$$

Injecting the expression of p^* into the equation $\boldsymbol{\theta}^* = u(\boldsymbol{\theta}^*, p^*)$, one obtains,

$$u\left(\boldsymbol{\theta}^*, \frac{\mathbf{1}^\top\boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right) - \boldsymbol{\theta}^* = \mathbf{0}.$$

Which is equivalent to,

$$(\mathbf{I} - \operatorname{diag}(\boldsymbol{\theta}^*))(\mathbf{I} + \operatorname{diag}(\boldsymbol{\theta}^*))\times \underbrace{\left(\beta s_p\left(\bar{p} - \frac{\mathbf{1}^\top\boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right)\mathbf{1} + (1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}_\theta(\boldsymbol{\theta}^*) - \boldsymbol{\theta}^*\right)}_{=\boldsymbol{g}(\boldsymbol{\theta}^*)-\boldsymbol{\theta}^*} = \mathbf{0}.$$

Thus, the equilibria $\boldsymbol{x}^* = (\boldsymbol{\theta}^*, p^*)$ are the point such that $p^* = \mathbf{1}^\top\boldsymbol{E}(\boldsymbol{\theta}^*)/\gamma$ and $\boldsymbol{\theta}^*$ verify

$$\begin{aligned}&(\mathbf{I} - \operatorname{diag}(\boldsymbol{\theta}^*))(\mathbf{I} + \operatorname{diag}(\boldsymbol{\theta}^*))(\boldsymbol{g}(\boldsymbol{\theta}^*) - \boldsymbol{\theta}^*) = \mathbf{0}\\ \Leftrightarrow\ &\forall i\in\mathcal{V},\ \ e_i^\top(\mathbf{I} - \operatorname{diag}(\boldsymbol{\theta}^*))(\mathbf{I} + \operatorname{diag}(\boldsymbol{\theta}^*))(\boldsymbol{g}(\boldsymbol{\theta}^*) - \boldsymbol{\theta}^*) = 0\\ \Leftrightarrow\ &\forall i\in\mathcal{V},\ \ (1-\theta_i)(1+\theta_i)\,e_i^\top(\boldsymbol{g}(\boldsymbol{\theta}^*) - \boldsymbol{\theta}^*) = 0\\ \Leftrightarrow\ &\forall i\in\mathcal{V},\ \ (1-\theta_i)(1+\theta_i)\left(e_i^\top\boldsymbol{g}(\boldsymbol{\theta}) - \theta_i\right) = 0,\end{aligned}$$

which is equivalent to $\boldsymbol{\theta}^*\in\mathcal{F}$. □

Proposition 1 provides a condition that must be satisfied by equilibria of the system (3) and, by extension, the set of all equilibria of the dynamics (3).

Remark 2. *It is noteworthy that* $\mathrm{Fix}(\boldsymbol{g}) \subset \mathcal{F}$. *Indeed, to ensure the existence of a nontrivial equilibrium (neither* $-\mathbf{1}$ *nor* $\mathbf{1}$*), it is sufficient to ensure that* $\mathrm{Fix}(\boldsymbol{g}) \neq \emptyset$.

The following lemma ensures the existence of fixed points for $\boldsymbol{g}$ under signal continuity, thereby establishing the existence of a nontrivial equilibrium for the dynamics (3).

Lemma 2. *Let* s_p *and* s_θ *be continuous functions. Then, the function* $\boldsymbol{g} : \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ *from* (4) *has at least one fixed point on* $\boldsymbol{\Theta}$.

Let us note that $\boldsymbol{\Theta}$ is convex and compact (recall that $\boldsymbol{\Theta} = [-1, 1]^N$) and $\boldsymbol{g} : \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ is continuous. Then, this lemma is a direct consequence of Brouwer's fixed point theorem [23].

Keeping in mind that a classical solution that received a lot of attention in opinion dynamics is the consensus or synchronization of opinions, it is natural to consider this particular situation in the sequel.

Let us note $\tilde{s}_p(\theta) : \Theta \to \Theta$, the function such that $\tilde{s}_p(\theta) = s_p\left(\bar{p} - NE\left(\theta\right)/\gamma\right)$.

Lemma 3. *Let* $\theta^* \in \mathrm{Fix}\left(s_\theta\right) \cap \mathrm{Fix}\left(\tilde{s}_p\right)$. *Then, the vector* $\boldsymbol{x}^* = (\boldsymbol{\theta}^*, \mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)/\gamma)$ *where* $\boldsymbol{\theta}^* = \theta^* \mathbf{1}$ *is a nontrivial equilibrium of* (3). *Moreover, if* s_p *is continuous, then* $\left|\mathrm{Fix}\left(s_\theta\right) \cap \mathrm{Fix}\left(\tilde{s}_p\right)\right| \leq 1$.

Proof. By definition, $E : \Theta \to \mathbb{R}$ and $s_p : \mathbb{R} \to \Theta$ are non-decreasing in their argument, yielding that $_p(\theta) = s_p\left(\bar{p} - NE\left(\theta\right)/\gamma\right)$ is a non-increasing function in θ. Since $\tilde{s}_p : \Theta \to \Theta$ is decreasing, it has at most one fixed point. Indeed by contradiction, assuming that $x, y \in \mathrm{Fix}\left(\tilde{s}_p\right)$ such that $x < y$, then one has $x = \tilde{s}_p(x) > \tilde{s}_p(y) = y$ which is a contradiction. Moreover, adding the assumption that $\tilde{s}_p$ is continuous and using Brouwer's fixed-point theorem [23] yields the existence of a fixed point. Thus, the fixed point of $\tilde{s}_p$ is unique under continuity assumption and one has $\left|\mathrm{Fix}\left(s_\theta\right) \cap \mathrm{Fix}\left(\tilde{s}_p\right)\right| \leq \left|\mathrm{Fix}\left(\tilde{s}_p\right)\right| = 1$.

Assuming $\mathrm{Fix}\left(s_\theta\right) \cap \mathrm{Fix}\left(\tilde{s}_p\right)$ being non-empty, let $\theta \in \mathrm{Fix}\left(s_\theta\right) \cap \mathrm{Fix}\left(\tilde{s}_p\right)$ and let $\boldsymbol{\theta} = \theta \mathbf{1}$. From Proposition 1, for $\boldsymbol{\theta}$ to be an equilibrium, it needs to verify (5). If $\theta \in \{-1, 1\}$ this condition is immediately met. But it is noteworthy that with the Standing Assumption 1 such points cannot be fixed point of $\tilde{s}_p$ and thus $-1, 1 \notin \mathrm{Fix}\left(s_\theta\right) \cap \mathrm{Fix}\left(\tilde{s}_p\right)$. If instead, $\theta \in (-1, 1)$, $\boldsymbol{\theta}$ has to be a fixed point of $\boldsymbol{g}$ and rewriting $\boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{\theta}$ yields

$$\begin{aligned}
&\boldsymbol{g}\left(\boldsymbol{\theta}\right) - \boldsymbol{\theta} = \beta\left(\tilde{s}_p\left(\boldsymbol{\theta}\right) - \boldsymbol{\theta}\right) + (1-\beta)\left(\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}_\theta\left(\boldsymbol{\theta}\right) - \theta\mathbf{1}\right) \\
&= \beta\left(\theta - \theta\right)\mathbf{1} + (1-\beta)\,\theta\boldsymbol{D}^{-1}\left(\boldsymbol{A}\mathbf{1} - \boldsymbol{D}\mathbf{1}\right) \\
&= -(1-\beta)\,\theta\boldsymbol{D}^{-1}\boldsymbol{L}\mathbf{1}.
\end{aligned}$$

But the all-ones vector $\mathbf{1}$ is an eigenvector of the Laplacian matrix with eigenvalue 0, one has $-(1-\beta)\,\theta\boldsymbol{D}^{-1}\boldsymbol{L}\mathbf{1} = \mathbf{0}$. □

Keeping in mind the particular case where the signals of opinions between neighbors are the opinions themselves, we can formulate the following remark.

Remark 3. *If* $s_\theta = \mathrm{Id}$, *lemma 3 states that* $\theta^* \in \mathrm{Fix}\left(\tilde{s}_p\right)$ *is a sufficient condition for* $\boldsymbol{x}^* = (\theta^*\mathbf{1}, NE\left(\theta^*\right)/\gamma)$ *to be an equilibrium of dynamics* (3).

It is then natural to ask if $\theta \in \mathrm{Fix}\left(s_\theta\right) \cap \mathrm{Fix}\left(\tilde{s}_p\right)$ is a necessary condition for consensus to be an equilibrium of dynamics (3). Let us note, the scalar version of $\boldsymbol{g}$ as $g : \Theta \to \Theta$ such that $g = \beta\tilde{s}_p + (1-\beta)\,s_\theta$. The following proposition reply to this question by proving that, indeed, the consensus is only attained at a fixed point of g.

Proposition 2. *Let* $\boldsymbol{x}^* = (\boldsymbol{\theta}^*, \mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)/\gamma)$ *be an equilibrium of* (3) *such that* $\exists \theta^* \in (-1, 1)$ *and* $\boldsymbol{\theta}^* = \theta^*\mathbf{1}$. *Then* $\theta^* \in \mathrm{Fix}\left(g\right)$.

Proof. We have the following equivalences

$$\begin{aligned}
&\boldsymbol{g}\left(\boldsymbol{\theta}^*\right) = \boldsymbol{\theta}^* \\
&\Leftrightarrow \beta s_p\left(\bar{p} - \frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right)\mathbf{1} + (1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}_\theta\left(\boldsymbol{\theta}^*\right) = \boldsymbol{\theta}^* \\
&\Leftrightarrow \beta s_p\left(\bar{p} - \frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right)\mathbf{1} + (1-\beta)\,\boldsymbol{s}_\theta\left(\boldsymbol{\theta}^*\right) + \\
&\qquad (1-\beta)\left(\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}_\theta\left(\boldsymbol{\theta}^*\right) - \boldsymbol{s}_\theta\left(\boldsymbol{\theta}^*\right)\right) = \boldsymbol{\theta}^* \\
&\Leftrightarrow \beta s_p\left(\bar{p} - \frac{NE(\theta^*)}{\gamma}\right)\mathbf{1} + (1-\beta)\,s_\theta\left(\theta^*\right)\mathbf{1} \\
&\qquad - (1-\beta)\,s_\theta\left(\theta^*\right)\boldsymbol{D}^{-1}\boldsymbol{L}\mathbf{1} = \theta^*\mathbf{1} \\
&\Leftrightarrow \beta s_p\left(\bar{p} - \frac{NE(\theta^*)}{\gamma}\right) + (1-\beta)\,s_\theta\left(\theta^*\right) = \theta^* \\
&\Leftrightarrow g\left(\theta^*\right) = \theta^* \Leftrightarrow \theta^* \in \mathrm{Fix}\left(\beta\tilde{s}_p + (1-\beta)\,s_\theta\right).
\end{aligned}$$

□

Now, if $s_\theta = \mathrm{Id}$, combining Remark 3 and Proposition 2 yields that consensus must be at fixed points of $\tilde{s}_p$, i.e. the vector $\boldsymbol{x}^* = (\theta^*\mathbf{1}, NE\left(\theta^*\right)/\gamma)$ is a equilibrium of dynamics (3) if and only if $\theta^* \in \mathrm{Fix}\left(\tilde{s}_p\right)$. The next Proposition goes further by stating that dynamics (3), in case of linear signaling dynamics over a connected graph, has a unique equilibrium which is a consensus.

Proposition 3. *Let* $\mathcal{G}$ *be a strongly connected graph and* $s_\theta = \mathrm{Id}$. *Then dynamics* (3) *has a unique equilibrium without extreme opinions given by* $\boldsymbol{x}^* = (\theta^*\mathbf{1}, NE\left(\theta^*\right)/\gamma)$ *such that* $\theta^* \in \mathrm{Fix}\left(\tilde{s}_p\right)$.

Proof. As stated above, combining Lemma 3 and Proposition 2, we know that $\boldsymbol{x}^*$ is one equilibrium of (3). To prove it is unique, let assume that $\boldsymbol{\theta}^*$ is a fixed point of $\boldsymbol{g}$, then

$$\begin{aligned}
\boldsymbol{g}\left(\boldsymbol{\theta}^*\right) = \boldsymbol{\theta}^* &\Leftrightarrow \beta s_p\left(\bar{p} - \frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right)\mathbf{1} + (1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{\theta}^* = \boldsymbol{\theta}^* \\
&\Leftrightarrow s_p\left(\bar{p} - \frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right)\mathbf{1} - \frac{1-\beta}{\beta}\boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{\theta}^* = \boldsymbol{\theta}^*.
\end{aligned}$$

In [24], it is shown that for any strongly connected graph, the associated Laplacian matrix $\boldsymbol{L}$ has eigenvalue 0 with multiplicity 1. Taking $\boldsymbol{v}$ the left eigenvector of $\boldsymbol{L}$ associated with the eigenvalue 0. Multiplying by $\boldsymbol{v}^\top\boldsymbol{D}$, on the left side the equation

above yields,

$$\begin{aligned}
&\boldsymbol{v}^\top \boldsymbol{D}\left(s_p\left(\bar{p}-\frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right)\mathbf{1}-\frac{1-\beta}{\beta}\boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{\theta}^*\right)=\boldsymbol{v}^\top \boldsymbol{D}\boldsymbol{\theta}^*\\
\Leftrightarrow\ & s_p\left(\bar{p}-\frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right)\boldsymbol{v}^\top \boldsymbol{D}\mathbf{1}-\frac{1-\beta}{\beta}\boldsymbol{v}^\top \boldsymbol{D}\boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{\theta}^*=\boldsymbol{v}^\top \boldsymbol{D}\boldsymbol{\theta}^*\\
\Leftrightarrow\ & s_p\left(\bar{p}-\frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right)=\frac{\boldsymbol{v}^\top \boldsymbol{D}\boldsymbol{\theta}^*}{\boldsymbol{v}^\top \boldsymbol{D}\mathbf{1}}.
\end{aligned}$$

This allows us to rewrite,

$$\begin{aligned}
\boldsymbol{g}(\boldsymbol{\theta}^*)=\boldsymbol{\theta}^* &\Leftrightarrow \beta\frac{\boldsymbol{v}^\top \boldsymbol{D}\boldsymbol{\theta}^*}{\boldsymbol{v}^\top \boldsymbol{D}\mathbf{1}}\mathbf{1}+(1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{\theta}^*=\boldsymbol{\theta}^*\\
&\Leftrightarrow \beta\frac{\mathbf{1}\boldsymbol{v}^\top \boldsymbol{D}}{\boldsymbol{v}^\top \boldsymbol{D}\mathbf{1}}\boldsymbol{\theta}^*+(1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{\theta}^*=\boldsymbol{\theta}^*\\
&\Leftrightarrow \left[\mathbf{I}-\left(\beta\frac{\mathbf{1}\boldsymbol{v}^\top \boldsymbol{D}}{\boldsymbol{v}^\top \boldsymbol{D}\mathbf{1}}+(1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\right)\right]\boldsymbol{\theta}^*=\mathbf{0}.
\end{aligned}$$

This linear map can now be seen as the Laplacian matrix of a new strongly connected, positively weighted graph. Indeed, the matrix $\mathbf{1}\boldsymbol{v}^\top \boldsymbol{D}/\boldsymbol{v}^\top \boldsymbol{D}\mathbf{1}$ can be interpreted as the adjacency matrix of a uniformly positively weighted all-to-all graph, with self-loops for each agent. Moreover, performing a convex combination of this new graph with the strongly connected, positively weighted influence graph $\boldsymbol{D}^{-1}\boldsymbol{A}$ results in a strongly connected graph. Finally, the degree matrix of this new graph is the identity matrix $\mathbf{I}$ since, $\forall i \in \mathcal{V}$, one has

$$e_i^\top\left(\beta\frac{\mathbf{1}\boldsymbol{v}^\top \boldsymbol{D}}{\boldsymbol{v}^\top \boldsymbol{D}\mathbf{1}}+(1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\right)=\beta+1-\beta=1.$$

Since the resulting mapping is a Laplacian matrix of a strongly connected graph, it is well known that its kernel is reduced to $\{\alpha\mathbf{1} \mid \alpha\in\mathbb{R}\}$. Therefore, the only fixed point of $\boldsymbol{g}$ is $\boldsymbol{\theta}^* \in \{\alpha\mathbf{1} \mid \alpha\in\mathbb{R}\}\cap \mathrm{Fix}(\tilde{s}_p)$, i.e., $\boldsymbol{\theta}^*=\theta^*\mathbf{1}$ such that $\theta^*\in\mathrm{Fix}(\tilde{s}_p)$. □

We note that Proposition 3 highlights a situation in which all the opinions reach a common value. It basically says that if each individual has access to the exact opinions of its neighbors in a strongly connected graph, they will reach a consensus. Moreover, the consensus value is determined by the external signal coupled with the opinion dynamics.

From now on, we denote by $g_i(\boldsymbol{\theta}) := e_i^\top \boldsymbol{g}(\boldsymbol{\theta})$ for any $i\in\mathcal{V}$. We note that equilibrium may depend not only on the signal function but also on the topology of the graph. Particularly, when symmetry of observation between neighbors arises, we can formulate the following result.

Proposition 4. *Let $\boldsymbol{x}^*=(\boldsymbol{\theta}^*,p^*)$ be a nontrivial equilibrium point of the dynamic* (3) *and let i, $j\in\mathcal{V}$ such that $i\neq j$ and either: "$\mathcal{N}_i=\mathcal{N}_j$" or "$i\in\mathcal{N}_j$, $j\in\mathcal{N}_i$ and $\mathcal{N}_j\setminus\{i\}=\mathcal{N}_i\setminus\{j\}$". Then we have $\theta_i^*=\theta_j^*$.*

Proof. If $\boldsymbol{x}^*$ is a nontrivial equilibrium of (3), one has $\boldsymbol{g}(\boldsymbol{\theta}^*)=\boldsymbol{\theta}^*$ yielding, for i and $j\in\mathcal{V}$,

$$\begin{cases} g_i(\boldsymbol{\theta}^*)-\theta_i^*=0\\ g_j(\boldsymbol{\theta}^*)-\theta_j^*=0\end{cases}\Leftrightarrow\begin{cases} g_i(\boldsymbol{\theta}^*)-\theta_i^*=0\\ g_i(\boldsymbol{\theta}^*)-\theta_i^*=g_j(\boldsymbol{\theta}^*)-\theta_j^*\end{cases}.$$

The second equation in the system above is rewritten as:

$$\frac{(1-\beta)}{n}\left(e_j^\top \boldsymbol{A}-e_i^\top \boldsymbol{A}\right)\boldsymbol{s}_\theta(\boldsymbol{\theta}^*)=\theta_j^*-\theta_i^*,$$

where $n=n_i=n_j$ the numbers of neighbors of i and j.

Let us note that $\forall k\in\mathcal{V}$, $e_k^\top \boldsymbol{A}$ is the vector with the component related to the neighbors of k equal 1 and all the others equal 0. If $\mathcal{N}_i=\mathcal{N}_j$, we have $e_i^\top \boldsymbol{A}=e_j^\top \boldsymbol{A}$ and the result became obvious. If i and j are neighbors, i.e. $j\in\mathcal{N}_i$ and $i\in\mathcal{N}_j$, one has $\mathcal{N}_i\neq\mathcal{N}_j$ but $\mathcal{N}_i\cap\mathcal{N}_j=\mathcal{N}_i\setminus\{j\}=\mathcal{N}_j\setminus\{i\}$. In this case, $e_i^\top \boldsymbol{A}-e_j^\top=e_j^\top \boldsymbol{A}-e_i^\top$. We proceed by contradiction and assume that $\theta_i^*\neq\theta_j^*$, without loss of generality, we can assume that $\theta_i<\theta_j$. Now, the equation becomes

$$\begin{aligned}
&\frac{(1-\beta)}{n}\left(e_i^\top-e_j^\top\right)\boldsymbol{s}_\theta(\boldsymbol{\theta}^*)=\theta_j^*-\theta_i^*\\
&\Leftrightarrow \frac{(1-\beta)}{n}\left(s_\theta(\theta_i^*)-s_\theta\left(\theta_j^*\right)\right)=\theta_j^*-\theta_i^*.
\end{aligned}$$

Since s_θ is monotonically increasing we obtain a contradiction. Indeed, the left-hand side is non-positive while the right-hand side is strictly positive. This is a contradiction, which implies that we must have $\theta_i^*=\theta_j^*$. □

Proposition 4 says that, at equilibrium, and for any monotonous increasing signaling function, two different agents with identical observation will have the same opinions. A direct consequence of this proposition is that if the graph is all-to-all, then the nontrivial equilibrium point corresponds to consensus.

3.2. Stability analysis of equilibria

In the following, we will assume that s_θ, s_p, and E are differentiable functions. From Proposition 1, the Jacobian of $\mathbf{F}$ is given by

$$\mathbf{J}(\boldsymbol{\theta},p)=\begin{bmatrix}\dfrac{\partial \boldsymbol{u}(\boldsymbol{\theta},p)}{\partial\boldsymbol{\theta}} & -\beta s_p'(\bar{p}-p)(\mathbf{I}-\mathrm{diag}(\boldsymbol{\theta})^2)\mathbf{1}\\ \nabla\boldsymbol{E}(\boldsymbol{\theta})^\top & 1-\gamma\end{bmatrix},\qquad(6)$$

with

$$\begin{aligned}
\frac{\partial \boldsymbol{u}(\boldsymbol{\theta},p)}{\partial\boldsymbol{\theta}}=\ &\mathbf{I}-2\mathrm{diag}(\boldsymbol{\theta})\,\mathrm{diag}(\boldsymbol{f}(\boldsymbol{\theta},p)-\boldsymbol{\theta})+\\
&\left(\mathbf{I}-\mathrm{diag}(\boldsymbol{\theta})^2\right)\left[(1-\beta)\,\boldsymbol{D}^{-1}\boldsymbol{A}\,\mathrm{diag}(\boldsymbol{s}_\theta'(\boldsymbol{\theta}))-\mathbf{I}\right].
\end{aligned}$$

We have seen earlier that some equilibria of the dynamics can contain extreme values in opinion, i.e. for such equilibrium it exists $i\in\mathcal{V}$, $\theta_i\in\{-1,1\}$. The following proposition states that such equilibria are unstable.

Theorem 1. *Let $\boldsymbol{x}^*=(\boldsymbol{\theta}^*,p^*)$ be an equilibrium point of the dynamics* (3)*, $\mathcal{G}$ be a strongly connected graph and $\boldsymbol{s}_\theta$, E and s_p be differentiable functions in $\boldsymbol{\theta}^*$ and $\bar{p}-p^*$, respectively. Then, if it exists $i\in\mathcal{V}$ such that $\theta_i^*\in\{-1,1\}$, then $\boldsymbol{x}^*$ is an unstable equilibrium.*

Proof. Let $\boldsymbol{x}^* = (\boldsymbol{\theta}^*, p^*)$ be an equilibrium of (3) such that there exists $i \in \mathcal{V}$ with $\theta_i \in \{-1, 1\}$. By permutation of the index, we can assume without loss of generality that $i = 1$, i.e. $\theta_1^* \in \{-1, 1\}$. Isolating the opinion of agent 1 in the Jacobian matrix given by (6), one can rewrite this Jacobian as the following block matrix

$$\mathbf{J}(\boldsymbol{\theta}, p) = \begin{bmatrix} \mathbf{J}_{11}(\boldsymbol{\theta}, p) & \mathbf{J}_{12}(\boldsymbol{\theta}, p) \\ \mathbf{J}_{21}(\boldsymbol{\theta}, p) & \mathbf{J}_{22}(\boldsymbol{\theta}, p) \end{bmatrix},$$

with $\mathbf{J}_{11} \in \mathbb{R}$ being scalar, $\mathbf{J}_{12}$, $\mathbf{J}_{21}^\top \in \mathbb{R}^N$ being N-vector and $\mathbf{J}_{22} \in \mathbb{R}^{N\times N}$ being a N-square matrix. Noting that $\theta_1^* \in \{-1, 1\}$, one obtains

$$\mathbf{J}_{11}(\boldsymbol{\theta}^*, \frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}) = 1 + 2\theta_1^{*2} - 2\theta_1^* g_1(\boldsymbol{\theta}^*) = 3 - 2\theta_1^* g_1(\boldsymbol{\theta}^*).$$

Moreover, non-diagonal elements of $\partial \boldsymbol{u}(\boldsymbol{\theta}, p)/\partial \boldsymbol{\theta}$ are weighted by $\mathbf{I} - \mathrm{diag}\,(\boldsymbol{\theta}^*)^2$ and, since $\theta_1^* \in \{-1, 1\}$, the first row of (6) is null except on its first coordinate yielding,

$$\mathbf{J}_{12}(\boldsymbol{\theta}^*, \frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}) = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} = \mathbf{0}_N{}^\top.$$

By the properties of the determinant of block-triangular matrices, the characteristic polynomial $P_{\mathbf{J}}$ of $\mathbf{J}$ is given by

$$P_{\mathbf{J}}(X) = \det(\mathbf{I}_{N+1}X - \mathbf{J}) = \det(X - \mathbf{J}_{11}) \det(\mathbf{I}_N X - \mathbf{J}_{22}).$$

Then, $\lambda = \mathbf{J}_{11}(\boldsymbol{\theta}^*, \mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)/\gamma) = 3 - 2\theta_1^* g_1(\boldsymbol{\theta}^*)$ is one of the eigenvalues of the Jacobian matrix (6). Now, either $\boldsymbol{\theta}^* = \pm\mathbf{1}$ or there exists $j \in \mathcal{V}$ such that $\theta_1^* \neq \theta_j^*$. In the first case, one has

$$\left| e_1^\top \boldsymbol{D}^{-1} \boldsymbol{A} \boldsymbol{s}_\theta(\boldsymbol{\theta}^*) \right| = \left| \frac{1}{n_1} \sum_{j\in\mathcal{N}_1} s_\theta\left(\theta_j^*\right) \right| = 1.$$

Then, for $\boldsymbol{\theta}^* = -\mathbf{1}$,

$$\lambda = 3 + 2\left(\beta \tilde{s}_p(-1) - (1-\beta)\right) = 1 + 2\beta\left(1 + \tilde{s}_p(-1)\right).$$

From Standing Assumption 1, we have $\bar{p} > Ne_{\min}/\gamma$ and since $s_p(p) = 0$ only if $p = 0$, one has $\tilde{s}_p(-1) > 0$. Moreover, noting that $0 < \beta$, yields,

$$\lambda = 1 + 2\beta\left(1 + \tilde{s}_p(-1)\right) > 1.$$

Then, if $\boldsymbol{\theta} = -\mathbf{1}$, at least one eigenvalues of (6) evaluated at $\boldsymbol{x}^*$ is strictly greater than 1. Then, by the linearisation principle, $\boldsymbol{x}^*$ is unstable.

If $\boldsymbol{\theta}^* = \mathbf{1}$, similar reasoning apply noting that, since $\bar{p} < Ne_{\max}/\gamma$, one has $\tilde{s}_p(1) = s_p(\bar{p} - Ne_{\max}/\gamma) < 0$ and then

$$\lambda = 3 - 2\left(\beta \tilde{s}_p(1) + (1-\beta)\right) = 1 + 2\beta\left(1 - \tilde{s}_p(1)\right) > 1.$$

Again, the linearisation principle implies that, if $\boldsymbol{\theta}^* = \mathbf{1}$ then $\boldsymbol{x}^*$ is unstable.

If $\boldsymbol{\theta}^*$ is neither $-\mathbf{1}$ nor $\mathbf{1}$, there exists $j \in \mathcal{V}$ such that $\theta_j \neq \theta_1$. Without loss of generality let $j \in \mathcal{N}_1$. Indeed, since the graph is connected and $\boldsymbol{\theta}^* \neq \pm\mathbf{1}$, there exists at least one $i \in \mathcal{V}$ with $\theta_i^* \in \{-1, 1\}$ such that there $\exists j \in \mathcal{N}_i$ verifying $\theta_j^* \neq \theta_i^*$. By permutation of the labeling, one can assume there $\exists j \in \mathcal{N}_1$ such that $\theta_j^* \neq \theta_1^*$. Now, one has,

$$\left| e_1^\top \boldsymbol{D}^{-1} \boldsymbol{A} \boldsymbol{s}_\theta(\boldsymbol{\theta}^*) \right| = \left| \frac{1}{n_1} \sum_{j\in\mathcal{N}_1} s_\theta\left(\theta_j^*\right) \right| < 1,$$

and since $0 < \beta < 1$, we have

$$\begin{aligned} |g_1(\boldsymbol{\theta}^*)| &= \left| \beta s_p\left(\bar{p} - \frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right) + (1-\beta)\, e_1^\top \boldsymbol{D}^{-1} \boldsymbol{A} \boldsymbol{s}_\theta(\boldsymbol{\theta}^*) \right| \\ &\leq \beta \left| s_p\left(\bar{p} - \frac{\mathbf{1}^\top \boldsymbol{E}(\boldsymbol{\theta}^*)}{\gamma}\right) \right| + (1-\beta) \left| e_1^\top \boldsymbol{D}^{-1} \boldsymbol{A} \boldsymbol{s}_\theta(\boldsymbol{\theta}^*) \right| < 1. \end{aligned}$$

We have then,

$$\lambda = 3 - 2\theta_1^* g_1(\boldsymbol{\theta}^*) \geq 3 - 2|g_1(\boldsymbol{\theta}^*)| > 1.$$

To conclude, if $\exists i \in \mathcal{V}$ such that $\theta_i^* \in \{-1, 1\}$, the Jacobian evaluated in this point $\mathbf{J}(\boldsymbol{\theta}^*, p^*)$ has at least one eigenvalue of modulus strictly greater than 1 and thus this equilibrium is unstable. □

The theorem states that, under Assumption 1, for a strongly connected graph, every equilibrium where at least one agent holds an extreme opinion is unstable. This implies that if opinions are allowed to spread freely, perturbing agents with extreme opinions will cause them to doubt and shift toward a less extreme stance.

It is important to emphasize that this result requires the assumption of strong connectivity in the graph. For example, if the graph $\mathcal{G}$ has an isolated disconnected agent, an equilibrium in which all the other (connected) agents reach consensus at some non-extreme opinion with the single agent holding the opposing extreme opinion can be locally asymptotically stable.

4. Analysis of the Fully Synchronized Opinions (FSO) case

In this section, we consider that the opinion dynamics evolves much faster than the air pollution state. Consequently, in the sequel the following condition holds true.

Assumption 1. *There exists a time-scale separation between the fast variation of θ and the slow variation of p.*

Based on the Assumption 1 we can use the standard approach in [25] to decouple the overall dynamics into the reduced order model and the boundary layer one. We recall that the boundary layer system evolves on the fast time scale and assumes that the slow variable is constant. In this case, under some regularity condition of signal s_θ, the opinions will converge to a consensus as proven in the next result.

Proposition 5. *Under Assumption 1, let s_θ be a k-Lipschitz function with $k < 1/(1-\beta)$. Then, for any $p \in \mathbb{R}_{\geq 0}$ the function $\boldsymbol{\theta} \mapsto \boldsymbol{f}(\boldsymbol{\theta}, p)$ has a unique fixed point in $\boldsymbol{\Theta}$ given by $\theta^*\mathbf{1}$ with $\theta^* = \mathrm{Fix}\left(\theta \mapsto \beta s_p(\bar{p} - p) + (1-\beta)\, s_\theta(\theta)\right)$.*

Proof. Let s_θ be a k-Lipschitz function, then for any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$ and any $p \in \mathbb{R}_{\geq 0}$, one has

$$\left\|\boldsymbol{f}(\boldsymbol{\theta}, p) - \boldsymbol{f}(\boldsymbol{\theta}', p)\right\|_2 \leq k(1-\beta)\left\|\boldsymbol{\theta} - \boldsymbol{\theta}'\right\|_2. \tag{7}$$

Indeed, by expanding the formulation of $\boldsymbol{f}$ in this specific case, one obtains

$$\begin{aligned}
&\left\|\boldsymbol{f}(\boldsymbol{\theta}, p) - \boldsymbol{f}(\boldsymbol{\theta}', p)\right\|_2 \\
&= \left\|\beta\left(s_p(\bar{p}-p) - s_p(\bar{p}-p)\right)\mathbf{1} + (1-\beta)\boldsymbol{D}^{-1}\boldsymbol{A}(\boldsymbol{s}_\theta(\boldsymbol{\theta}) - \boldsymbol{s}_\theta(\boldsymbol{\theta}'))\right\|_2 \\
&= \left\|(1-\beta)\boldsymbol{D}^{-1}\boldsymbol{A}\left(\boldsymbol{s}_\theta(\boldsymbol{\theta}) - \boldsymbol{s}_\theta(\boldsymbol{\theta}')\right)\right\|_2 \\
&\leq (1-\beta)\left\|\boldsymbol{D}^{-1}\boldsymbol{A}\right\|_{2,2}\left\|\boldsymbol{s}_\theta(\boldsymbol{\theta}) - \boldsymbol{s}_\theta(\boldsymbol{\theta}')\right\|_2 \leq (1-\beta)k\left\|\boldsymbol{\theta} - \boldsymbol{\theta}'\right\|_2.
\end{aligned}$$

Then, for a constant $p \in \mathbb{R}_{\geq 0}$, $\boldsymbol{f}$ is a contraction with respect to its first variable $\boldsymbol{\theta}$. Therefore, by Banach fixed-point theorem [23], $\boldsymbol{\theta} \mapsto \boldsymbol{f}(\boldsymbol{\theta}, p)$ has a unique fixed point. Let $\boldsymbol{\theta}^*$ be this fixed point, it must verify

$$\begin{aligned}
&\boldsymbol{f}(\boldsymbol{\theta}^*, p) = \boldsymbol{\theta}^* \Leftrightarrow \beta s_p(\bar{p}-p)\mathbf{1} + (1-\beta)\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}_\theta(\boldsymbol{\theta}^*) = \boldsymbol{\theta}^* \\
&\Leftrightarrow \beta s_p(\bar{p}-p)\mathbf{1} + (1-\beta)\boldsymbol{s}_\theta(\boldsymbol{\theta}^*) - \boldsymbol{\theta}^* \\
&\qquad + (1-\beta)\left(\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}_\theta(\boldsymbol{\theta}^*) - \boldsymbol{s}_\theta(\boldsymbol{\theta}^*)\right) = \mathbf{0} \\
&\Leftrightarrow \beta s_p(\bar{p}-p)\mathbf{1} + (1-\beta)\boldsymbol{s}_\theta(\boldsymbol{\theta}^*) - \boldsymbol{\theta}^* - (1-\beta)\boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{s}_\theta(\boldsymbol{\theta}^*) = \mathbf{0}.
\end{aligned}$$

But now, notice that $\tilde{f} : \theta \mapsto \beta s_p(\bar{p}-p) + (1-\beta)s_\theta(\theta)$ is also a continuous contraction of Θ. Then, as before, by Banach fixed-point theorem, it has a unique fixed point, say θ^*. Moreover, taking $\boldsymbol{\theta} = \theta^*\mathbf{1}$ verify the above equation since $\boldsymbol{L}\mathbf{1} = \mathbf{0}$. □

Remark 4. *The condition $k < 1/(1-\beta)$ imposes a sufficiently high influence of the pollution dynamics on the opinion dynamics compared with the variation speed of the signaling function s_θ.*

The time-scale separation and the emergence of consensus shown in Proposition 5 motivate us to focus on the analysis of the reduced order system under the assumption that all opinions in the network are synchronized. From now on, s_θ, s_p and E are differentiable functions and all the agents share the same opinion.

Definition 1. *We say that an opinion state $\boldsymbol{\theta}$ is in the FSO regime if and only if $\forall i, j \in \mathcal{V}$, $\theta_i = \theta_j$. i.e. $\boldsymbol{\theta} = \theta\mathbf{1}$ for $\theta \in \Theta$.*

The following proposition studies the effect of the dynamics once FSO has been reached.

Proposition 6. *The FSO manifold*

$$M = \{\boldsymbol{x} = (\boldsymbol{\theta}, p) \mid \boldsymbol{\theta} = \theta\mathbf{1}, \text{ for } \theta \in [-1,1],\ p \in \mathbb{R}_{\geq 0}\} \subset \mathbb{R}^{N+1},$$

is forward invariant for the dynamics (3)*. i.e. $\forall \boldsymbol{x} \in M$, $\boldsymbol{F}(\boldsymbol{x}) \in M$.*

Proof. Let $\boldsymbol{x} \in M$, by applying the dynamics (3) to $\boldsymbol{x} = (\theta\mathbf{1}, p)$ with $\theta \in [-1,1]$ and $p \in \mathbb{R}_{\geq 0}$, one can write

$$\mathbf{F}(\boldsymbol{x}) = \begin{pmatrix}\boldsymbol{u}(\theta\mathbf{1}, p)\\ v(\theta\mathbf{1}, p)\end{pmatrix} = \begin{pmatrix}\theta\mathbf{1} + \left(\mathbf{I} - \operatorname{diag}(\theta\mathbf{1})^2\right)(\boldsymbol{f}(\theta\mathbf{1}, p) - \theta\mathbf{1})\\ (1-\gamma)p + \mathbf{1}^\top\boldsymbol{E}(\theta\mathbf{1})\end{pmatrix}.$$

For any, $\theta \in [-1,1]$, $\mathbf{1}^\top\boldsymbol{E}(\theta\mathbf{1}) = NE(\theta)$, and then $(1-\gamma)p + NE(\theta) \in \mathbb{R}_{\geq 0}$ since $\gamma < 1$ and $E(\theta) \geq 0$. Moreover, developing

$$\begin{aligned}
&\theta\mathbf{1} + \left(\mathbf{I} - \operatorname{diag}(\theta\mathbf{1})^2\right)(\boldsymbol{f}(\theta\mathbf{1}, p) - \theta\mathbf{1}) \\
&= \theta\mathbf{1} + (1-\theta^2)\mathbf{I}\left(\beta s_p(\bar{p}-p)\mathbf{1} + (1-\beta)\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}_\theta(\theta\mathbf{1}) - \theta\mathbf{1}\right) \\
&= \theta\mathbf{1} + (1-\theta^2)\mathbf{I}\left(\beta s_p(\bar{p}-p)\mathbf{1} + (1-\beta)s_\theta(\theta)\boldsymbol{D}^{-1}\boldsymbol{A}\mathbf{1} - \theta\mathbf{1}\right) \\
&= \theta\mathbf{1} + (1-\theta^2)\left(\beta s_p(\bar{p}-p)\mathbf{1} + (1-\beta)s_\theta(\theta)\mathbf{1} - \theta\mathbf{1}\right) \\
&= \left(\theta + (1-\theta^2)\left(\beta s_p(\bar{p}-p) + (1-\beta)s_\theta(\theta) - \theta\right)\right)\mathbf{1} := \theta'\mathbf{1}.
\end{aligned}$$

And since $\theta' \in [-1,1]$, one obtains $\boldsymbol{x} \in M$. □

To simplify the presentation, in the FSO regime, we will denote $\theta(k)$ and the common opinion. In other words, we omit the agent index when referring to its opinion or emission.

Under the assumption that $\boldsymbol{\theta}$ is in FSO regime, the opinion dynamics of any agent coupled with the environment in (3) rewrites as

$$\boldsymbol{x}(k+1) = \mathbf{F}(\boldsymbol{x}(k)) = \begin{pmatrix}u(\theta(k), p(k))\\ v(\theta(k), p(k))\end{pmatrix}, \tag{8}$$

with the scalar function

$$u(\theta, p) = \theta + \left(1-\theta^2\right)(f(\theta, p) - \theta),$$

where

$$f(\theta, p) = \beta s_p(\bar{p}-p) + (1-\beta)s_\theta(\theta).$$

Now, the Jacobian of the dynamics becomes:

$$\mathbf{J}(\theta, p) = \begin{bmatrix}\dfrac{\partial u}{\partial\theta}(\theta, p) & -\beta\left(1-\theta^2\right)s_p'(\bar{p}-p)\\ E'(\theta) & 1-\gamma\end{bmatrix}$$

with

$$\frac{\partial u}{\partial\theta}(\theta, p) = 1 - 2\theta(f(\theta, p) - \theta) + \left(1-\theta^2\right)((1-\beta)s_\theta'(\theta) - 1).$$

The dynamics preserves the properties established in the general case. For instance, the equilibrium $\boldsymbol{x}^* = (\theta^*, p^*)$ with $\theta^* \in \{-1, 1\}$ is unstable by Proposition 1. This justifies our focus on the stability analysis of equilibrium of the form $\boldsymbol{x}^* = (\theta^*, NE(\theta^*)/\gamma)$, with $\theta^* \in \operatorname{Fix}(g)$, (recall that $g = \beta\tilde{s}_p + (1-\beta)s_\theta$). In FSO, the Jacobian slightly changes due to the term $\partial u/\partial\theta$ that becomes:

$$\frac{\partial u}{\partial\theta}(\theta, p) = 1 + \left(1-\theta^2\right)((1-\beta)s_\theta'(\theta) - 1).$$

Now, the characteristic polynomial of $\mathbf{J}$ is given by:

$$\begin{aligned}
P(X) = X^2 &- \left(2 + \left(1-\theta^2\right)((1-\beta)s_\theta'(\theta) - 1) - \gamma\right)X \\
&+ \left(1 + \left(1-\theta^2\right)((1-\beta)s_\theta'(\theta) - 1)\right)(1-\gamma) \\
&+ \beta\left(1-\theta^2\right)E'(\theta)s_p'\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right).
\end{aligned} \tag{9}$$

And its discriminant is given by:

$$\Delta_P = \left(\left(1-\theta^2\right)\left((1-\beta)\, s'_\theta(\theta) - 1\right) + \gamma\right)^2 - 4\beta\left(1-\theta^2\right) E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right).$$

To simplify the stability analysis of equilibria, let us state the following intermediate lemma that provides conditions to ensure that roots of a polynomial are of modulus less than 1.

Lemma 4. *Let $P(X) = X^2 + bX + c$ be a second-order polynomial with $b, c \in \mathbb{R}$ such that its discriminant $\Delta_P = b^2 - 4c \geq 0$ and denote its roots $\lambda_1 = \left(-b - \sqrt{\Delta_P}\right)/2$ and $\lambda_2 = \left(-b + \sqrt{\Delta_P}\right)/2$. Then, $|\lambda_1| < 1$ and $|\lambda_2| < 1$ if and only if $|b| < 1 + c$ with $b \in (-2, 2)$.*

Proof. Note that $|\lambda_1| < 1$ and $|\lambda_2| < 1$ is equivalent with

$$-1 < \lambda_1 \leq \lambda_2 < 1 \Leftrightarrow -2 + b < -\sqrt{\Delta_P} \leq \sqrt{\Delta_P} < 2 + b$$

yielding $0 \leq \sqrt{\Delta_P} < 2 - |b|$ which requires that $b \in (-2, 2)$. Moreover

$$\begin{aligned}\sqrt{\Delta_P} < 2 - |b| &\Leftrightarrow b^2 - 4c < 4 - 4|b| + b^2 \\ &\Leftrightarrow |b| < 1 + c\end{aligned}$$

which concludes the proof. □

The following proposition characterizes the stability of the equilibrium point $\theta^* \in \mathrm{Fix}(g) \cap (-1, 1)$ with respect to the parameters β, γ and the function s_θ and s_p.

Theorem 2. *Let $x^* = (\theta^*, p^*)$ be an equilibrium of the dynamics (8) with $\theta^* \in \mathrm{Fix}(g) \cap (-1, 1)$. Let P be the characteristic polynomial of $\mathbf{J}$ and note Δ_P its discriminant. Then, x^* is locally exponentially stable if and only if one of the following is verified:*

- $\Delta_P \geq 0$ *and*

$$\max\left(1 - \frac{\gamma + \left(1-\theta^{*2}\right)}{\left(1-\theta^{*2}\right) s'_\theta(\theta^*)}, \frac{\gamma\left(s'_\theta(\theta^*) - 1\right)}{\gamma s'_\theta(\theta^*) + E'(\theta^*) s'_p\left(\bar{p} - \frac{NE(\theta^*)}{\gamma}\right)}\right) < \beta.$$

- $\Delta_P < 0$ *and*

$$\beta\left(E'(\theta^*) s'_p\left(\bar{p} - \frac{NE(\theta^*)}{\gamma}\right) - (1-\gamma)\, s'_\theta(\theta^*)\right) < \frac{\gamma}{\left(1-\theta^{*2}\right)} - \left(s'_\theta(\theta^*) - 1\right)(1-\gamma).$$

Proof. Notice that in the FSO setup, dynamics (3) reduces to a two-dimensional nonlinear dynamical system for which the study of its Jacobian's eigenvalues is sufficient to characterize the local stability of an equilibrium. Those eigenvalues are the roots of the polynomial (9) i.e., the roots of $P(x) = X^2 + bX + c$ with

$$b = -\left(2 + \left(1-\theta^2\right)\left((1-\beta)\, s'_\theta(\theta) - 1\right) - \gamma\right)$$

and

$$c = 1 - \gamma + \left(1-\theta^2\right)\left((1-\beta)\, s'_\theta(\theta) - 1\right)(1-\gamma) + \beta\left(1-\theta^2\right) E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right).$$

In the case where, $\Delta_P \geq 0$, we can use Lemma 4. Therefore, $|\lambda_1| < 1$ and $|\lambda_2| < 1$ is equivalent with $b \in (-2, 2)$ and $|b| < 1 + c$. First, $b \in (-2, 2)$ yields

$$\begin{aligned}&-2 < b < 2 \\ \Leftrightarrow\ &-2 < -\left(2 + \left(1-\theta^2\right)\left((1-\beta)\, s'_\theta(\theta) - 1\right) - \gamma\right) < 2 \\ \Leftrightarrow\ &-4 + \gamma < \left(1-\theta^2\right)\left((1-\beta)\, s'_\theta(\theta) - 1\right) < \gamma \\ \Leftrightarrow\ &1 + \frac{\gamma - 4}{\left(1-\theta^2\right)} < (1-\beta)\, s'_\theta(\theta) < 1 + \frac{\gamma}{\left(1-\theta^2\right)} \\ \Leftrightarrow\ &\frac{1}{s'_\theta(\theta)} + \frac{\gamma - 4}{\left(1-\theta^2\right) s'_\theta(\theta)} < (1-\beta) < \frac{1}{s'_\theta(\theta)} + \frac{\gamma}{\left(1-\theta^2\right) s'_\theta(\theta)} \\ \Leftrightarrow\ &1 - \frac{\gamma + \left(1-\theta^2\right)}{\left(1-\theta^2\right) s'_\theta(\theta)} < \beta < 1 + \frac{4 - \gamma - \left(1-\theta^2\right)}{\left(1-\theta^2\right) s'_\theta(\theta)}. \qquad (10)\end{aligned}$$

Let's observe that $0 < 4 - \gamma - \left(1-\theta^2\right), \forall\ \gamma \in (0,1),\ \theta \in \Theta$. This implies that $1 + \frac{4 - \gamma - \left(1-\theta^2\right)}{\left(1-\theta^2\right) s'_\theta(\theta)} > 1$. On the other hand, $\beta \in (0, 1)$ which simplifies (10) as $1 - \frac{\gamma + \left(1-\theta^2\right)}{\left(1-\theta^2\right) s'_\theta(\theta)} < \beta \leq 1$. It is noteworthy that $2 - \gamma - \left(1-\theta^2\right) > 0,\ \forall\ \gamma \in (0,1),\ \theta \in \Theta$. Thus, $\beta < 1 + \frac{2 - \gamma - \left(1-\theta^2\right)}{\left(1-\theta^2\right) s'_\theta(\theta)} \Leftrightarrow$

$$\beta\left(1-\theta^2\right) s'_\theta(\theta) < 2 + \left(1-\theta^2\right)\left(s'_\theta(\theta) - 1\right) - \gamma \Leftrightarrow b < 0.$$

Consequently,

$$\begin{aligned}&|b| < 1 + c \Leftrightarrow -b < 1 + c \\ \Leftrightarrow\ &2 - \gamma + \left(1-\theta^2\right)\left((1-\beta)\, s'_\theta(\theta) - 1\right) < 2 - \gamma + \\ &\left(1-\theta^2\right)\left[\left((1-\beta)\, s'_\theta(\theta) - 1\right)(1-\gamma) + \beta E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right)\right] \\ \Leftrightarrow\ &\gamma\left((1-\beta)\, s'_\theta(\theta) - 1\right) < \beta E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right) \\ \Leftrightarrow\ &\gamma\left(s'_\theta(\theta) - 1\right) < \beta\left(\gamma s'_\theta(\theta) + E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right)\right) \\ \Leftrightarrow\ &\frac{\gamma\left(s'_\theta(\theta) - 1\right)}{\gamma s'_\theta(\theta) + E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right)} < \beta.\end{aligned}$$

Combining the inequality above with the first inequality in (10) one gets that if $\Delta_P \geq 0$ then $|\lambda_1| < 1$ and $|\lambda_2| < 1$ is equivalent with

$$\max\left(1 - \frac{\gamma + \left(1-\theta^2\right)}{\left(1-\theta^2\right) s'_\theta(\theta)}, \frac{\gamma\left(s'_\theta(\theta) - 1\right)}{\gamma s'_\theta(\theta) + E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right)}\right) < \beta.$$

Let us assume now that $\Delta_P < 0$. The complex roots of P have the same modulus. Let λ_C be the eigenvalue defined by $\lambda_C = \dfrac{-b + i\sqrt{-\Delta_P}}{2}$. Then $|\lambda_C| = \dfrac{b^2 - b^2 + 4c}{4} = c$ and the following equivalences hold

$$
\begin{aligned}
c < 1 \Leftrightarrow\ & \left(1-\theta^2\right)\left((1-\beta)\, s'_\theta(\theta) - 1\right)(1-\gamma) \\
& + \beta\left(1-\theta^2\right) E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right) < \gamma \\
\Leftrightarrow\ & \left((1-\beta)\, s'_\theta(\theta) - 1\right)(1-\gamma) \\
& + \beta E'(\theta) + s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right) < \frac{\gamma}{(1-\theta^2)} \\
\Leftrightarrow\ & \left(s'_\theta(\theta) - 1\right)(1-\gamma) \\
& + \beta\left(E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right) - (1-\gamma)\, s'_\theta(\theta)\right) < \frac{\gamma}{(1-\theta^2)} \\
\Leftrightarrow\ & \beta\left(E'(\theta) s'_p\left(\bar{p} - \frac{NE(\theta)}{\gamma}\right) - (1-\gamma)\, s'_\theta(\theta)\right) \\
& < \frac{\gamma}{(1-\theta^2)} - \left(s'_\theta(\theta) - 1\right)(1-\gamma)\,.
\end{aligned}
$$

This concludes the proof. □

It is noteworthy that Lemma 1 states that the set $\mathcal{X} = \Theta \times [Ne_{\min}/\gamma, Ne_{\max}/\gamma]$ is forward invariant for (3) and then for (8). Thus, when the local stability conditions from Theorem 2 are not satisfied, the trajectory will exhibit either convergence to a limit cycle or a strange attractor.

Let us notice that β characterizes the weight of the air pollution signal in the update of the opinions. When $\beta \to 1$, the opinions will mostly be influenced by the environment, and the peers' influence is negligible. Therefore, when the pollution disappears at high speed (γ small), the dynamics will reach a stable equilibrium. If γ is high enough one obtains a discrete-time prey-predator-like behavior with permanent oscillations.

Let us consider the next two particular cases of the signal s_θ.

4.1. Linear signals

Simpler conditions can be exhibited in the linear signaling OD case and are given in the following corollary.

Corollary 1. *Let $s_\theta(\theta) = \theta$, for all $\theta \in \Theta$ and let $x^* = (\theta^*, p^*)$ be an equilibrium of the dynamics (8) with $\theta^* \in \mathrm{Fix}\,(g) \cap (-1, 1)$. Let P be the characteristic polynomial of $\mathbf{J}$ and note Δ_P its discriminant. Then, x^* is stable if and only if either $\Delta_P \geq 0$ or $\Delta_P < 0$ and*

$$
\beta\left(E'(\theta^*) s'_p\left(\bar{p} - \frac{NE(\theta^*)}{\gamma}\right) - (1-\gamma)\right) < \frac{\gamma}{(1-\theta^{*2})}. \tag{11}
$$

The proof is immediate from Proposition 2 by replacing $s'_\theta(\theta) = 1$, for all $\theta \in \Theta$.

These conditions allow us to exhibit the particular point where the topological nature of the equilibrium change and bifurcation may occur. Numerical illustrations of this behavior will be provided in the next section.

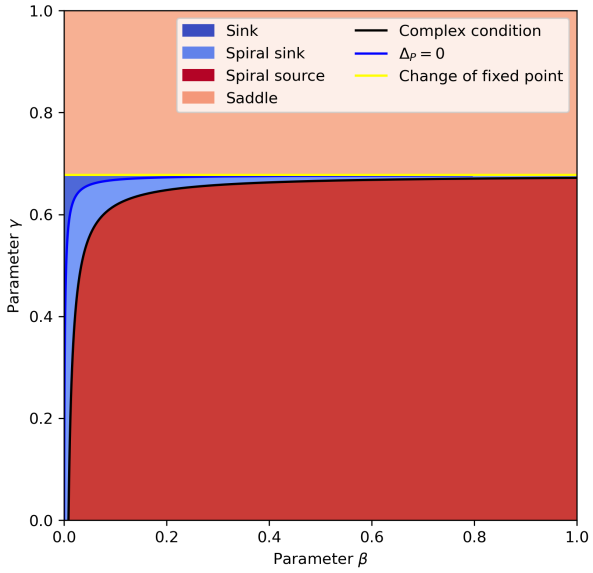


Figure 1: Topological classification of equilibrium $x^* = (\theta, NE(\theta)/\gamma)$ for the dynamics (8), where $\theta \in \mathrm{Fix}\,(g)$ with $s_\theta(\theta) = \theta$, $s_p = s$, as in (12) with $\delta = 0.05$, $E(\theta) = \theta + 1$ for $0 < \beta, \gamma < 1$, $N = 1$ and $\bar{p} = 2.5$.

4.2. Piece-wise linear signals

In this section, we will look at the particular case where the signal function s_θ and s_p are given by the function

$$
s(x) = \begin{cases} -1 & \text{if} \quad x \leq -\delta \\ \dfrac{x}{\delta} & \text{if} -\delta \leq x \leq \delta \\ 1 & \text{if} \quad x \geq \delta. \end{cases} \tag{12}
$$

parameterized by $\delta > 0$. This function is piece-wise linear and approximates the sgn function for values of δ close to 0. We will also make the assumption that $E(\theta) = \theta + 1$.

In the particular case of linear OD coupled with the piece-wise linear air pollution signal, i.e. $s_\theta(\theta) = \theta$, $\forall \theta \in \Theta$ and s_p as in (12), we know from Proposition 3 and Lemma 3 that there exists an unique fixed point of the function g. The fixed point of the function g in this case is given by the following expression

$$
l(\gamma, \bar{p}, \delta) = \begin{cases} \dfrac{\gamma\bar{p} - 1}{\gamma\delta + 1} & \text{if} \quad \gamma \leq \dfrac{2}{\bar{p} - \delta} \\ 1 & \text{otherwise.} \end{cases} \tag{13}
$$

The topological nature of the equilibrium associated is illustrated on the β, γ-plane on Figure 1 for some $\bar{p}$ fixed. We can observe that the nature of the equilibrium changes with respect to the parameters β and γ and is unstable for most of the parameter values. The complex condition black line is given by equation (11) and the yellow one from the condition of (13).

In the more general case when both s_θ, s_p are piece-wise linear as in (12), the number of equilibrium points varies with respect to the parameters and accordingly one has different expressions. The following lemma provides conditions under which certain points belong Fix(g). We recall here that $g(\theta) = \beta s(\bar{p} - (\theta + 1)/\gamma) + (1 - \beta)\, s(\theta)$.

Lemma 5. *Let s_θ and s_p be piece-wise linear as in* (12)*. The following statements hold true:*

- *if* $\max\{\theta, \bar{p} - E(\theta)/\gamma\} \leq -\delta$*, then* $-1 \in \mathrm{Fix}(g)$.
- *if* $\min\{\theta, \bar{p} - E(\theta)/\gamma\} \geq \delta$*, then* $1 \in \mathrm{Fix}(g)$.
- *if* $\theta \geq \delta$ *and* $\bar{p} - E(\theta)/\gamma \leq -\delta$*, then* $-1 + 2\beta \in \mathrm{Fix}(g)$.
- *if* $\theta \leq -\delta$ *and* $\bar{p} - E(\theta)/\gamma \geq \delta$*, then* $1 - 2\beta \in \mathrm{Fix}(g)$.
- *if* $|\theta| \geq \delta$ *and* $|\bar{p} - E(\theta)/\gamma| \leq \delta$*, then*
$$\frac{sgn(\theta)\,\delta\,(1-\beta) + \beta\,(\bar{p} - 1/\gamma)}{\delta + \beta/\gamma} \in \mathrm{Fix}(g). \tag{14}$$
- *if* $\max\{|\theta|, |\bar{p} - E(\theta)/\gamma|\} \leq \delta$*, then*
$$\frac{\beta\,(\bar{p} - 1/\gamma)}{\delta + \beta/\gamma - (1-\beta)} \in \mathrm{Fix}(g). \tag{15}$$
- *if* $|\theta| \leq \delta$ *and* $|\bar{p} - E(\theta)/\gamma| \geq \delta$*, then*
$$\frac{sgn(\bar{p} - E(\theta)/\gamma)\,\beta}{1 - (1-\beta)/\delta} \in \mathrm{Fix}(g). \tag{16}$$

The proof of this Lemma is straightforward, every item corresponds to a specific combination between the different possible linear behaviors of the signals s_θ and s_p in (12).

The topological classification of these fixed points can be numerically addressed as in the case of OD with a linear signaling. The numerical analysis is longer and we skip it here since one has to consider all the possible values in $\mathrm{Fix}(g)$.

5. Numerical Results

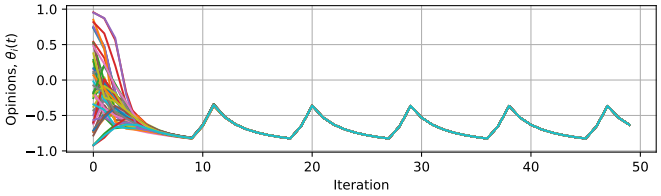
This section provides numerical illustrations of the coupled system's behavior. Note that we are numerically analyzing the behavior in the FSO scenario by illustrating the convergence to equilibrium, to a limit cycle, or chaotic trajectories. Additionally, we examine the non-synchronized case, highlighting the formation of clusters and emphasizing again the chaotic behavior. For the sake of simplicity, in the sequel, we take $E(\theta) = \theta + 1$ and denote by s the piece-wise linear approximation of the sign function defined in (12) with $\delta = 0.05$. Moreover, for the rest of the paper, we will also take s_p piece-wise linear, i.e., $s_p = s$.

5.1. Fully Synchronized Opinion

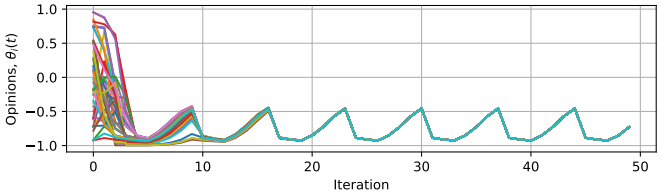
Let us notice that, synchronization occurs not only when the system reaches a fixed equilibrium as shown in Figure 2 for two types of signals. The rate of synchronization is influenced by the signal function and the parameter β: lower values of β result in faster synchronization, similar to linear opinion dynamics with exogenous input. Conversely, as β approaches 1, synchronization takes more iterations due to the greater influence of the environment over neighboring interactions, potentially preventing the synchronization.

In the FSO regime, all agents in the system share the same opinion, simplifying the model's dynamics to a two-dimensional system encompassing the common opinion θ and the air pollution state p.

We first examine scenarios with $s_\theta = \mathrm{Id}$ and later, we will illustrate similar behaviors with different interaction signals.



(a) $s_\theta = \mathrm{Id}$. Synchronization reached in 9 iterations.



(b) $s_\theta = s$, i.e. piece-wise linear. Synchronization reached in 16 iterations.

Figure 2: Evolution of $\boldsymbol{\theta}$ with two signal functions s_θ interacting over a common (randomly generated) strongly connected graph. Parameters: $\beta = 0.3$, $\gamma = 0.5$, $N = 50$, $p(0) = \bar{p} = 25$ and $\boldsymbol{\theta}(0)$ taken as i.i.d. random uniform variables on $[-1, 1]$.

Equilibrium and Limit cycle

Under the FSO assumption, the dynamics (8) can exhibit various behaviors depending on the signal and parameters. Notably, the dynamics can converge to an equilibrium or reach a limit cycle.

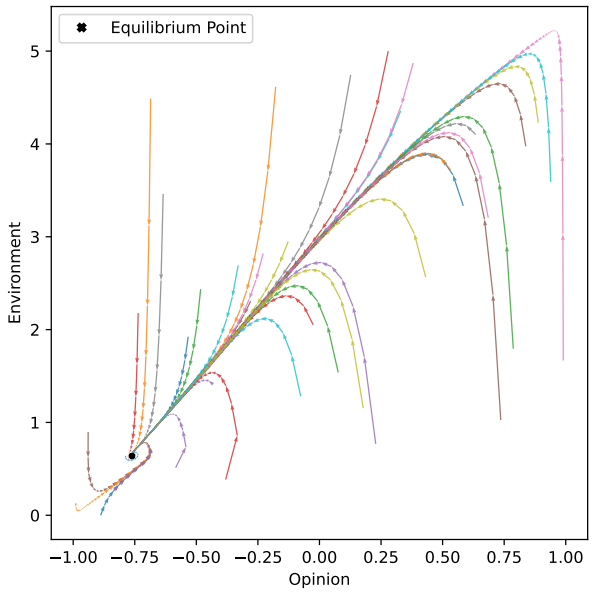


Figure 3: Phase portrait of dynamics (8) for several trajectories with random initial conditions. Trajectories converge to the unique nontrivial equilibrium. Parameters: $\beta = 0.04$, $\gamma = 0.375$ and $\bar{p} = 0.6$.

As illustrated in Figure 3, consistent with Proposition 3, trajectories converge to a unique, attractive in this case, equilibrium $x^* = (\theta^*, E(\theta^*)/\gamma)$, where $\theta^* \in \mathrm{Fix}\left(\tilde{s}_p\right)$.

Additionally, the dynamics can reach a limit cycle, as shown in Figure 4, where a 11-period limit cycle appears to stabilize. The length and stability of the limit cycle depend on the param-

eters β and γ as will be explained in the next section.

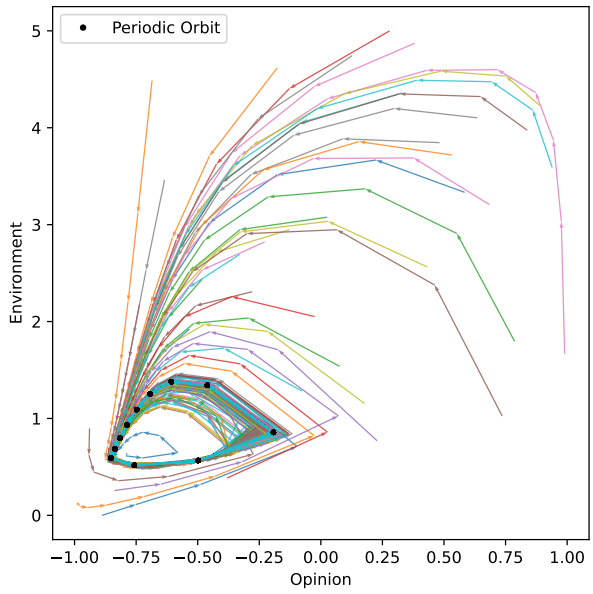


Figure 4: Phase portrait of dynamics (8) for several trajectories with random initial conditions. Trajectories converge to a 11-period limit cycle. Parameters: $\beta = 0.345$, $\gamma = 0.375$ and $\bar{p} = 0.6$.

The presence of limit cycles indicates a repeating pattern generated by short-term social memory. Population tends to take action when there is a peak of pollution but once the pollution decreases the population considers the problem is solved and returns to standard behavior. This suggests that without interventions to reward the ecologic behavior, the high pollution episodes will persist in time.

Chaos

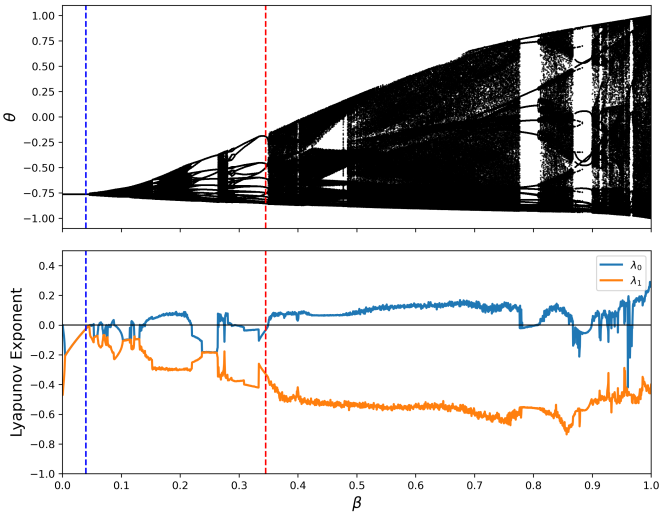


Figure 5: Bifurcation diagram for $0 < \beta < 1$ of the FSO regime opinion and Lyapunov exponents of the system. The system undergoes different behaviors: convergence to equilibrium, limit cycle, and chaos. The blue line represents the situation of Figure 3 and the red is the one of Figure 4. Parameters $\gamma = 0.375$ and $\bar{p} = 0.6$ and initial condition $\theta(0) = -0.5$ and $p(0) = 1.6$.

As evidenced in Figure 5, the FSO regime can also manifest chaotic behaviors. This figure presents the bifurcation diagram and the associated Lyapunov exponents.

We recall that a dynamical system with Lyapunov exponents that are less than 0 has trajectories that converge to each other. They synchronously reach a stable attractor like a fixed point or a limit cycle. On the other hand, positive Lyapunov exponents indicate chaos or instability, where trajectories diverge exponentially from each other. Under fixed values of γ and $\bar{p}$, a wide range of β values can lead to chaotic dynamics in (8).

Moreover, we can observe that the equilibrium presented in Figure 3 and the limit cycle of Figure 4 are both attractive since their associated Lyapunov exponents are negative. It is noteworthy that in Figure 5, limit cycles present on the bifurcation diagram are always associated with a negative Lyapunov exponent and thus stabilizing.

From an environmental perspective, chaotic behavior is similar to limit cycles. Indeed, although there is no periodic pattern in the evolution of opinions and air pollution, there are permanent oscillations between ecologic actions and high emissions. The main difference lies in the unpredictability of the public response and consequently of the environmental state.

Other signals

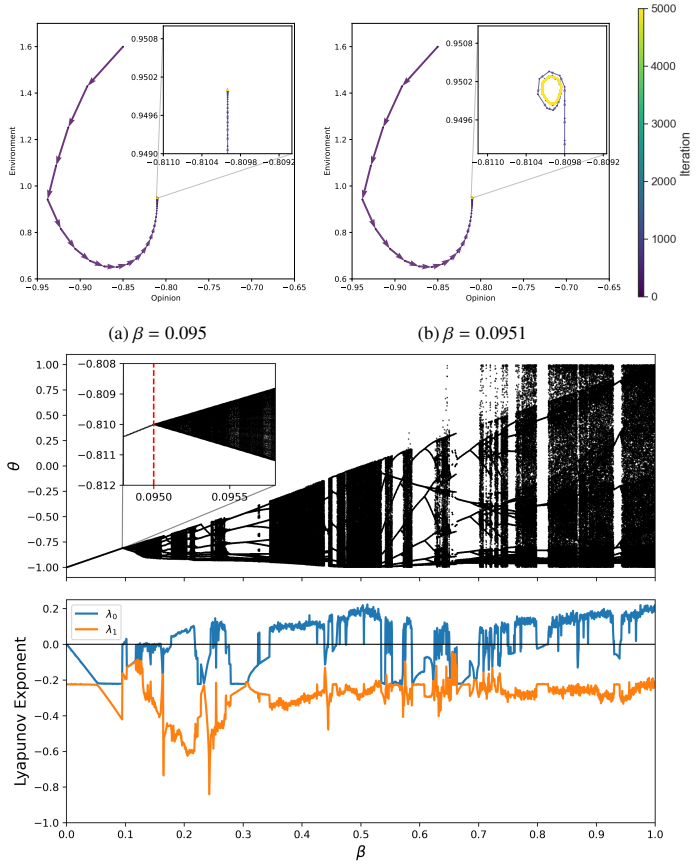
We have noticed through extensive numerical simulations that similar qualitative behaviors are preserved across various signaling functions. This robustness suggests that the fundamental characteristics of the model are independent of the signaling functions. Indeed, we can see in Figure 6 that trajectories can reach equilibrium for some parameter β like in Figure 6a or strange attractor in Figure 6b. Here we can observe the first bifurcation occurring at $\beta = 0.095$ for the associated parameters. This bifurcation, illustrated in Figure 6c, presents a transition from convergence toward a stable equilibrium to a strange attractor, which seems to be a Hopf bifurcation.

The presence of a strange attractor, as evidenced by the successive zooms on the phase portrait in Figure 7, suggests that the system's trajectories are governed by an invariant yet non-periodic underlying structure. During simulations, it has been observed that this attractor folds onto itself as the parameters β and γ change. At lower values of β and γ, as depicted in Figure 6, the attractor does not exhibit significant folding. However, as β approaches 1 and for $\gamma = 0.5$, the folding phenomenon becomes more pronounced, especially as γ approaches 0.25, as shown in Figure 7.

While the behavior in the FSO case can be mathematically analyzed, the complexity of the system with multiple interacting agents that are not synchronized hampers a theoretical study.

5.2. *Out of sync case*

In the following, we numerically analyze the case when the agents do not synchronize. This phenomenon is illustrated in Figure 8, where agents do not synchronize for a value of β close to 1, even after a large number of iterations. The lack of synchronization and the chaotic behavior are interdependent, and



(c) Bifurcation diagram of the opinion of the agent and Lyapunov exponents of the system.

Figure 6: Top: The first 5000 steps of trajectories of (8) before (a) and just after (b) the bifurcation point highlighted (in red) in the bifurcation diagram (c), i.e. $\beta = 0.095$ and $\beta = 0.0951$. A behavioral transition is observed, moving from an equilibrium to a strange attractor. We use $s_\theta = s$ and parameters $\bar{p} = 1$ and $\gamma = 0.2$, with initial condition $\theta(0) = -0.85$ and $p(0) = 1.6$. (c) provides the Bifurcation diagram and Lyaponuv exponents for $0 < \beta < 1$.

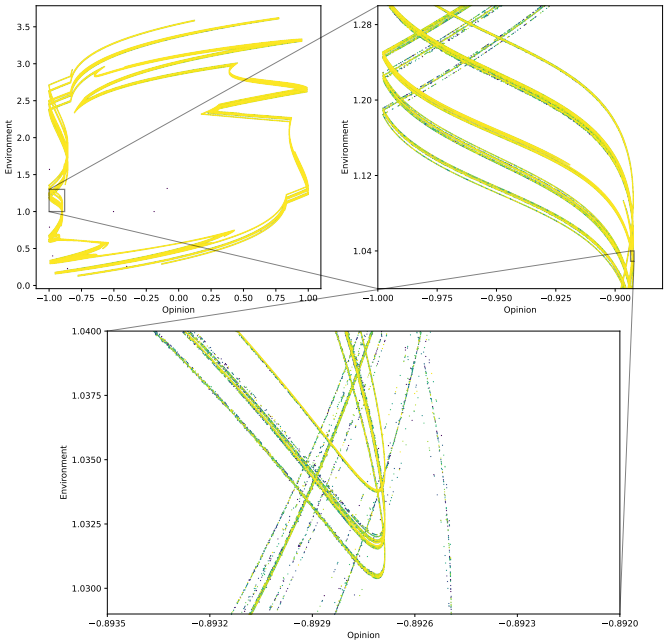


Figure 7: The trajectory from $k = 0$ to $k = 10,000,000$ with successive zoom for $s_\theta = s$ with parameters: $\beta = 0.916$, $\gamma = 0.5$, and initial condition $p(0) = \bar{p} = 1$ and $\theta(0) = -0.5$. We observe that this trajectory exhibits a fractal pattern.

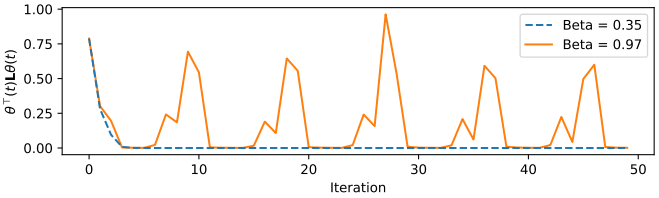


Figure 8: Chaotic de-synchronization highlighted by $\boldsymbol{\theta}^\top \boldsymbol{L}\boldsymbol{\theta}$ plotted over 50 steps over a randomly generated connected undirected graph for $s_\theta = \mathrm{Id}$. For $\beta = 0.35$, opinions reach FSO over a limit cycle in a few iterations, whereas, for $\beta = 0.97$ consensus is never reached (verified up to 10^6 steps). Parameters: $\gamma = 0.5$, $N = 50$, $p(0) = \bar{p} = 25$ and $\boldsymbol{\theta}(0)$ taken as i.i.d. random uniform variables on $[-1, 1]$.

we call this phenomenon chaotic de-synchronization as synchronization occurs in all other regimes.

An examination of the bifurcation diagrams for all agents, as shown in Figure 9, reveals varying degrees of synchronization. For $\beta < 0.95$, agents tend to synchronize towards FSO; however, for $\beta > 0.95$, agents diverge onto different trajectories. This divergence is further emphasized by the behavior of the associated Lyapunov exponents, which initially evolve in a similar manner but begin to diverge significantly beyond $\beta = 0.95$. This suggests that for values of β close to one, the chaotic behavior intensifies, potentially becoming too strong to permit synchronization.

While the absence of synchronization in chaotic linear signal dynamics is significant, it is not the only scenario in which agents fail to synchronize. Indeed, under certain conditions and signal configurations, clusters of aligned opinions can form, reflecting the graph topology's influence.

Polarized cluster formation

Let us now examine the formation of clusters of opinions under two scenarios, leveraging Proposition 4 to highlight the impact of graph topology, particularly its symmetry, on equilibrium states. We focus on two distinct graph topologies: a block stochastic graph (blocks represented by complete graphs that are connected through few edges) and a sparser 2D-lattice graph. These structures reveal that the formation and behavior of opinion of clusters (stability, polarization, periodicity) significantly depend on the initial conditions and the graph's structure. Specifically, we observe stable polarized clusters in block stochastic graphs as shown in Figure 10 for the former scenario.

On the other hand, Figure 11 provides insights into the dynamics on a 50×50 square lattice, where initial opinions are randomly distributed and evolve through communication with adjacent nodes. This simulation underscores the formation of resilient clusters that maintain their integrity over numerous iterations, converging to clusters on limit cycles, as can be seen from the opinion trajectory of the nodes in Figure 11.

The primary distinction between the two cases is that in the

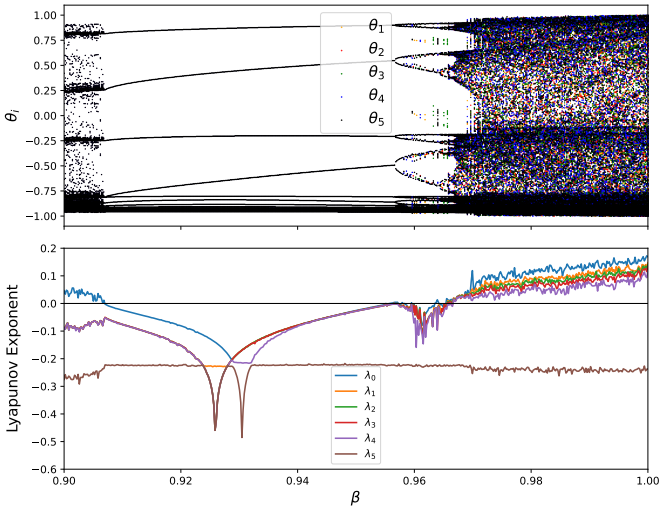
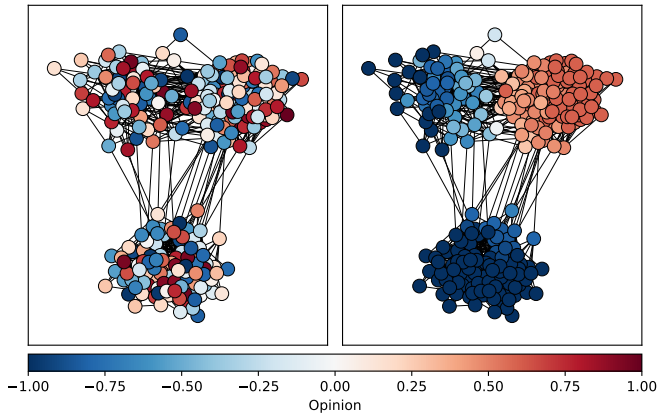
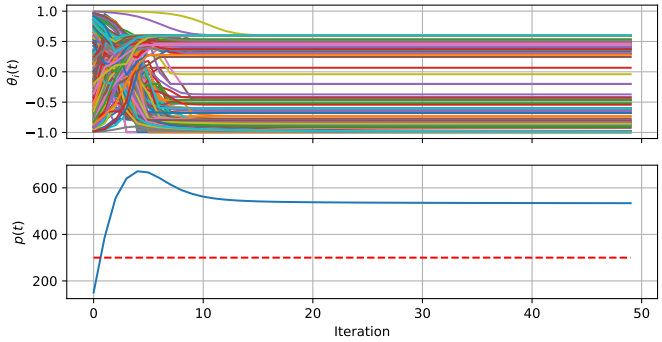


Figure 9: Bifurcation diagram and Lyapunov exponents for agents 5 over a fully connected graph and the system for $0.9 < \beta < 1$ for $s_\theta = \mathrm{Id}$. For $\beta < 0.95$, agents synchronize whereas for β close to 1, synchronization does not occur anymore. Parameters: $\gamma = 0.2$, $\bar{p} = 4.5$, $p(0) = 30$ and $\boldsymbol{\theta}(0)$ taken as i.i.d. random uniform variables on $[-1, 1]$.



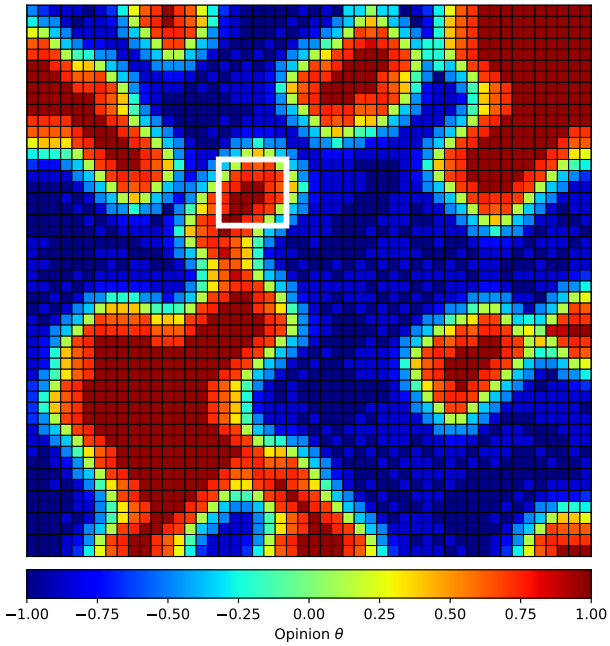
(a) Clustered graph. Left: initial conditions. Right: after 50 iterations.



(b) Evolution of $\boldsymbol{\theta}$ and environment over 50 iterations. Initial conditions: $p(0) = 175$ and $\boldsymbol{\theta}(0)$ taken as i.i.d. random uniform variables on $[-1, 1]$.

Figure 10: Evolution of the opinion on a randomly generated clustered graph for $s_\theta = s$. We observe that a polarized opinion cluster forms over the graph cluster. Parameters $\beta = 0.2$, $\gamma = 0.44$, $N = 300$ and $\bar{p} = 350$.

first, the environment reaches an equilibrium and then emits a constant signal, allowing the OD to converge to the stable



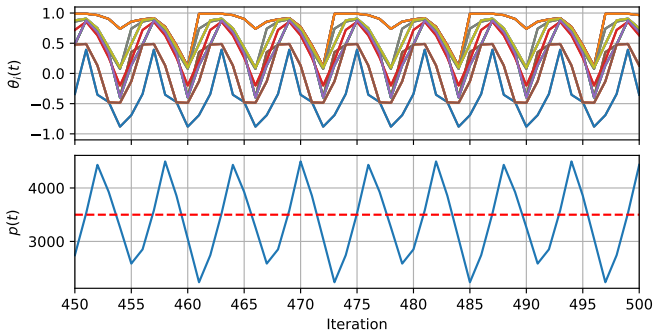


Figure 11: Visualization of Opinion Dynamics on a 50×50 square Lattice for $s_\theta = s$. The resultant opinions after 500 iterations are represented by each colored square cell. Agents engage in communication with their adjacent cells (above, below, left, and right). The first sub-figure represents opinions over the last 50 iterations of the nodes in the white square. We observe a periodic evolution that maintains the structure of the cluster. The sub-figure below presents the air pollution evolution over the last 50 iterations. Parameters: $\beta = 0.49$, $\gamma = 0.6$, $p(0) = \bar{p} = 3500$, $\boldsymbol{\theta}(0)$ taken as i.i.d. random uniform variables on $[-1, 1]$

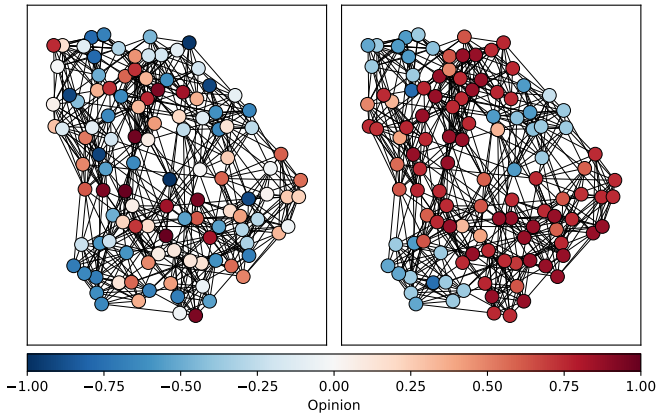
equilibrium shown in Figure 10. Conversely, in the second case, the structure of the clusters is preserved but is driven by the air pollution oscillating around $\bar{p}$. The formation of clusters typically occurs in a short amount of time, as depicted in Figure 10.

Small-world

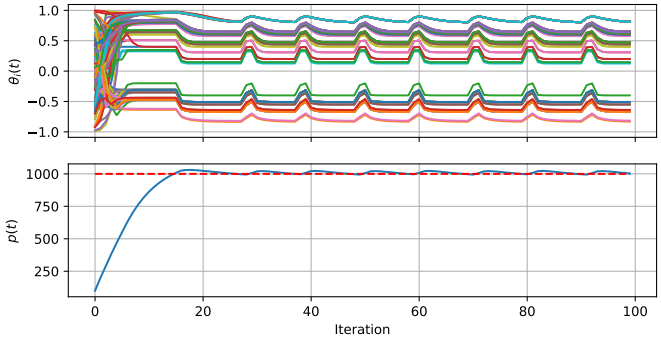
This section focuses on numerical simulations conducted on small-world networks, which are analogous to social networks due to their high clustering and short path lengths relative to random networks. Small-world properties are quantified using two metrics: the small-world coefficient σ and the small-world measure ω. A network is considered a small-world if $\sigma > 1$, and its classification between lattice-like ($\omega \approx 1$) and random-like ($\omega \approx -1$) structures is indicated by ω values around zero.

Additionally, the diameter of the graph, the longest of all the shortest paths between pairs of nodes, can provide further insight into network topology.

We analyze two graphs from the NetworkX Python library to demonstrate these concepts. The first graph represents the network of American football games between Division IA colleges during the regular season of Fall 2000 [26], and the second depicts a portion of a Facebook social network [27]. Our goal here is to demonstrate that the behaviors of dynamics (3) observed in the previous section are scalable and the numerical analysis can be extended to various types of graphs.



(a) Graph of American football clubs, where edges represent games between clubs. The graph contains 115 nodes and is a small-world network with $\sigma = 4.13$ and $\omega = -0.1$. Left: initial conditions. Right: after 100 iterations.



(b) Evolution of $\boldsymbol{\theta}$ and environment over 100 iterations. Initial conditions: $p(0) = 175$ and $\boldsymbol{\theta}(0)$ taken as i.i.d. random uniform variables on $[-1, 1]$.

Figure 12: Evolution of the opinion on the football club graph for $s_\theta = s$. We observe that polarized opinion clusters form even if the graph does not exhibit obvious clusters. Parameters $\beta = 0.2$, $\gamma = 0.2$, and $\bar{p} = 1000$.

Figure 12 highlights the polarization of opinions in two main clusters despite the absence of clear separation in the initial graph (see Figure 12a). This behavior mirrors that observed in regular grid graphs (see Figure 11), suggesting that graph regularity is not necessary for cluster formation; rather, it is influenced by parameters β and γ. We remark that the cluster formation is relatively fast and is achieved before the signal p reaches the threshold $\bar{p}$ as shown in Figure 12b.

In Figure 13 we consider a portion of a Facebook social network [27] and we plot the equilibrium reached in opinions. The behavior of the network heavily depends on the initial state of opinions. With the same parameters but a different initial

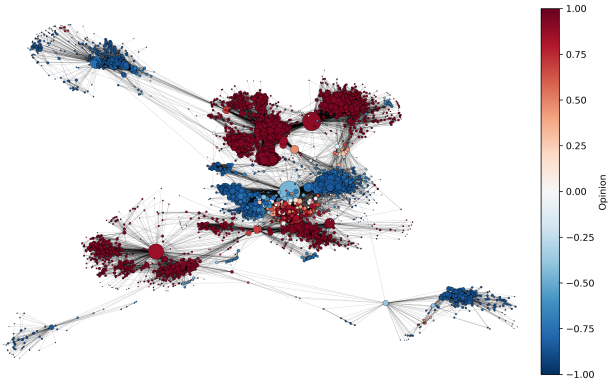


Figure 13: Facebook Network graph, representing anonymized users within ten friend lists. Edges correspond to friendships. The graph contains 4039 nodes and has a diameter of 8. For more details, see [**?**]. The displayed opinions are at equilibrium for $s_\theta = s_p = s$ with $\delta = 0.05$. Parameters: $\beta = 0.1$, $\gamma = 0.2$, and $\bar{p} = 25,000$. Initial conditions: $p(0) = 23,750$ and $\boldsymbol{\theta}(0)$ as i.i.d. random uniform variables on $[-1, 1]$, except for biased clusters at top-left and bottom-right around $\theta = -0.95$.

condition, the coupled system can reach a limit cycle or exhibit chaotic behavior (please check video limit cycle and video chaos). Notably, FSO is rarely observed in simulations, likely due to the topological clustering inherent to community-based friend networks.

These observations confirm that the behaviors noted previously can occur in diverse small-world network structures, not limited by the regularity or specific topology of the graphs.

Remark 5. *While not explicitly demonstrated in this section, it is important to note that these dynamical behaviors (FSO, convergence to one equilibrium, limit cycles, and chaotic patterns) are not unique to the specific signals used in the presented examples. Similar behaviors can be replicated with a variety of other signaling functions for both opinions and atmospheric pollution. This indicates that the underlying dynamics of the system are robust to changes in the way agents interact and respond to the air pollution dynamics, suggesting a broad applicability of our findings across different types of network-driven processes.*

6. Conclusion

In this paper, we have introduced and analyzed a nonlinear opinion dynamics model coupled with an external signal. Specifically, we consider that the external dynamics represent a very simple pollution model in which the emission level depends on the emissions of the individuals in the social network. Conversely, opinions are influenced by both the signal from their neighbors and the pollution level. We have shown that different behaviors are possible, ranging from convergence to a steady state to a chaotic behavior of the coupled dynamics. The chaotic regime is usually characterized by a strange attractor and numerical simulations demonstrate fractal patterns. Another interesting phenomenon observed is that of chaotic de-

synchronization, where agents that synchronize in limit cycles or equilibria get de-synchronized due to the chaotic behavior.

References

[1] E. Ising, Contribution to the theory of ferromagnetism, Ph.D. thesis, University of Hamburg (1924).

[2] K. Sznajd-Weron, J. Sznjad, Opinion evolution in closed community, International Journal of Modern Physics C 11 (6) (2000) 1157–1165.

[3] G. Deffuant, D. Neau, F. Amblard, G. Weisbuch, Mixing beliefs among interacting agents, Advances in Complex Systems 3 (2000) 87–98.

[4] R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence models, analysis, and simulation, Journal of Artificial Societies and Social Simulation 5 (3) (2002).

[5] N. E. Friedkin, E. C. Johnsen., Social influence and opinions., Journal of Mathematical Sociology. 15 (1990) 193–206.

[6] I.-C. Morărescu, A. Girard, Opinion dynamics with decaying confidence: Application to community detection in graphs, IEEE Transactions on Automatic Control 56 (8) (2011) 1862 – 1873.

[7] M. H. DeGroot, Reaching a consensus, Journal of the American Statistical Association 69 (345) (1974) 118–121.

[8] C. Altafini, Consensus problems on networks with antagonistic interactions, IEEE Transactions on Automatic Control 58 (4) (2013) 935–946.

[9] A. Martins, Continuous opinions and discrete actions in opinion dynamics problems, International Journal of Modern Physics C 19 (4) (2008) 617–625.

[10] N. R. Chowdhury, I.-C. Morărescu, S. Martin, S. Srikant, Continuous opinions and discrete actions in social networks: a multi-agent system approach, in: Proceedings 55th IEEE Conference on Decision and Control, 2016, pp. 1739–1744.

[11] F. Ceragioli, P. Frasca, Consensus and Disagreement: The Role of Quantized Behaviors in Opinion Dynamics, SIAM Journal on Control and Optimization 56 (2) (2018) 1058–1080.

[12] D. J. Hofmann, J. H. Butler, P. P. Tans, A new look at atmospheric carbon dioxide, Atmospheric Environment 43 (12) (2009) 2084–2086. doi:10.1016/j.atmosenv.2008.12.028.

[13] F. Joos, R. Roth, J. S. Fuglestvedt, G. P. Peters, I. G. Enting, W. von Bloh, et al., Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: A multi-model analysis, Atmospheric Chemistry and Physics 13 (5) (2013) 2793–2825. doi:10.5194/acp-13-2793-2013.

[14] N. Leach, S. Jenkins, Z. Nicholls, C. Smith, J. Lynch, M. Cain, et al., FaIRv2.0.0: A generalized impulse response model for climate uncertainty and future scenario exploration, Geoscientific Model Development 14 (2021) 3007–3036. doi:10.5194/gmd-14-3007-2021.

[15] A. Garg, A. Kumar, N. C. Gupta, Comprehensive study on impact assessment of lockdown on overall ambient air quality amid COVID-19 in Delhi and its NCR, India, Journal of Hazardous Materials Letters 2 (2021) 100010.

[16] A. Bizyaeva, A. Franci, N. E. Leonard, Nonlinear Opinion Dynamics With Tunable Sensitivity, IEEE Transactions on Automatic Control 68 (3) (2023) 1415–1430.

[17] Z. He, X. Lai, Bifurcation and chaotic behavior of a discrete-time predator–prey system, Nonlinear Analysis: Real World Applications 12 (1) (2011) 403–417.

[18] Y. Lin, Q. Din, M. Rafaqat, A. A. Elsadany, Y. Zeng, Dynamics and Chaos Control for a Discrete-Time Lotka-Volterra Model, IEEE Access (2020).

[19] J. S. Weitz, C. Eksin, K. Paarporn, S. P. Brown, W. C. Ratcliff, An oscillating tragedy of the commons in replicator dynamics with game-environment feedback, Proceedings of the National Academy of Sciences 113 (47) (2016) E7518–E7525. doi:10.1073/pnas.1604096113.

[20] A. R. Tilman, J. B. Plotkin, E. Akçay, Evolutionary games with environmental feedbacks, Nature Communications 11 (1) (2020) 915. doi:10.1038/s41467-020-14531-6.

[21] K. Frieswijk, L. Zino, M. Cao, A. S. Morse, Modeling the Co-evolution of Climate Impact and Population Behavior: A Mean-Field Analysis⋆, IFAC-PapersOnLine 56 (2) (2023) 7381–7386. doi:10.1016/j.ifacol.2023.10.355.

[22] A. Couthures, T. Mongaillard, V. S. Varma, S. Lasaulce, I.-C. Morărescu, Analysis of a continuous opinion and discrete action model coupled with an external dynamics, in: 2024 European Control Conference (ECC), 2024, pp. 3392–3397. doi:10.23919/ECC64448.2024.10591291.

[23] A. Granas, J. Dugundji, Fixed Point Theory, Springer Monographs in Mathematics, Springer New York, New York, NY, 2003.

[24] J. J. P. Veerman, R. Lyons, A Primer on Laplacian Dynamics in Directed Graphs (Feb. 2020). arXiv:2002.02605.

[25] P. Kokotović, H. Khalil, J. O'Reilly, Singular Perturbation Methods in Control: Analysis and Design, SIAM, 1999.

[26] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences 99 (12) (2002) 7821–7826.

[27] J. Leskovec, J. Mcauley, Learning to discover social circles in ego networks, in: F. Pereira, C. Burges, L. Bottou, K. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 25, Curran Associates, Inc., 2012.

Bifurcation analysis of an opinion dynamics model coupled with an environmental dynamics

Anthony Couthures[1], Anastasia Bizyaeva[2], Vineeth S. Varma[1,3], Alessio Franci[4], Irinel-Constantin Morărescu[1,3]

***Abstract*— We consider an opinion dynamics model coupled with an environmental dynamics. Based on a forward invariance argument, we can simplify the analysis of the asymptotic behavior to the case when all the opinions in the social network are synchronized. Our goal is to emphasize the role of the trust given to the environmental signal in the asymptotic behavior of the opinion dynamics and implicitly of the coupled system. To do that, we conduct a bifurcation analysis of the system around the origin when the trust parameter is varying. Specific conditions are presented for both pitchfork and Hopf bifurcation. Numerical illustration completes the theoretical findings.**

***Index Terms*— Bifurcation analysis, Opinion dynamics, Nonlinear systems**

I. INTRODUCTION

Modeling the dynamics of climate change and environmental processes is of pressing importance and has received significant attention over recent decades. It is noteworthy that opinion dynamics play an important role in the environmental processes and vice-versa. On one hand, individuals' opinions and behaviors are shaped by social interactions. Such interactions are often modeled via networked or multi-agent systems where consensus, polarization, and other collective phenomena emerge [1]–[4]. On the other hand, human actions can have a profound and sometimes irreversible impact on the environment. This duality is particularly evident in contexts such as climate change debates, sustainable behavior adoption, and collective decision-making in environmental policy [5], [6]. Although Opinion Dynamics (OD) and environmental processes have been extensively addressed separately, the interplay between the two remains insufficiently explored. Some research directions considering this interaction include the evolutionary game perspective [7], [8] and the dynamical systems one [9], [10].

In this paper, we propose and analyze a coupled model that integrates opinion dynamics with environmental feedback as a continuous-time extension of recent work [10]. Each agent in the network holds a continuously evolving opinion representing, for instance, a spectrum of attitudes from pro-environmental to anti-environmental behavior. Agents update their opinions based on two distinct influences: (i) social interactions with neighbors, mediated by a signal function, and (ii) the perceived state of the environment, which is itself affected by the collective behavior of the agents. The environmental state evolves according to a linear dynamic equation controlled by aggregate opinions, and agents indirectly perceive the environmental condition via a response function. This formulation captures the inherent feedback loop between individual behavior and the state of the environment.

The main contributions of this paper are summarized as follows. We consider a simplified model coupling the opinion and environment dynamics. Mathematically, the model is formulated as a system of ordinary differential equations in which both the agents' opinions and the environmental state evolve continuously over time. Our first technical result establishes the forward invariance of the synchronization manifold (all the opinions in the social network coincide). Next, under appropriate assumptions, we conduct a detailed analysis of the Fully Synchronized Opinion coupled with the Environment (FSOE) dynamics of the system, characterizing the conditions under which the system exhibits singular points. In the FSOE setup, we conduct a bifurcation analysis demonstrating both pitchfork and Hopf bifurcations emanating from the trivial equilibrium. In other words, we characterize the range of parameters guaranteeing the presence of oscillating behaviors between opinions and environment.

The rest of the paper is organized as follows. In Section II, we describe the coupled opinion–environment model along with the setup under consideration. Section III states a preliminary result on the forward invariance allowing to reduce the analysis to the FSOE case. Section IV is devoted to the analysis of the FSOE dynamics, including the characterization of equilibria, the identification of singular points, and the bifurcation analysis. Finally, Section V presents short numerical simulations that illustrate the system's behavior under different parameter settings. We conclude the paper in Section VI with a summary of the main results.

Notation We will denote by $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ the set of real and non-negative real numbers, respectively. For a vector $\boldsymbol{x} \in \mathbb{R}^N$, we denote by x_i the i-th component of $\boldsymbol{x}$. The components of the matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ are denoted a_{ij}. The i-th vector of the canonical basis of $\mathbb{R}^N$ is $\boldsymbol{e}_i$ and $\mathbf{1}$ is the vector of $\mathbb{R}^N$ with all components equal to 1.

The work of V.S. Varma and I.C. Morărescu was supported by project DECIDE, no. 57/14.11.2022 funded under the PNRR I8 scheme by the Romanian Ministry of Research, Innovation, and Digitisation

[1]Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France. anthony.couthures@univ-lorraine.fr

[3]associated with Automation Department, Technical University of Cluj-Napoca, Memorandumului 28, 400114 Cluj-Napoca, Romania.

[2]Sibley School of Mechanical and Aerospace Engineering at Cornell University

[4]Department of Electrical Engineering and Computer Science of the University of Liege and WEL Research Institute, Wavre, Belgium.

We also use the standard notation $\mathrm{diag}(\boldsymbol{x}) \in \mathbb{R}^{N\times N}$ for the diagonal matrix with diagonal elements given by the vector $\boldsymbol{x} \in \mathbb{R}^N$. For a function $f : \mathcal{X} \to \mathcal{X}$, we denote $\mathrm{Fix}(f) = \{x \in \mathcal{X} \mid f(x) = x\}$ the set of fixed points of f in $\mathcal{X}$.

II. Problem formulation

We consider the classical multi-agent framework in which N individuals/agents belonging to the set $\mathcal{V} = \{1, \dots, N\}$ interact according to an *undirected fixed graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. We denote by $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ its *adjacency matrix*, i.e., $a_{ij} = 1$ if $(i, j) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. We denote, by $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ its *degree matrix*, i.e., $\boldsymbol{D} = \mathrm{diag}\,(\boldsymbol{d})$ where $d_i = \sum_{j=1}^{N} a_{ij}$ for all $i \in \mathcal{V}$. The graph $\mathcal{G}$ is *connected* if one can find a path in the graph connecting any two different agents.

Since $\boldsymbol{D}^{-1}\boldsymbol{A}$ is a stochastic matrix, from the Perron-Frobenius theorem, we have the following result.

Lemma 1 (Perron-Frobenius): Let $\mathcal{G}$ be a connected graph. Then, the normalized adjacency matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$ has a simple eigenvalue 1, and all other eigenvalues have modulus strictly less than 1. Moreover, the vector $\mathbf{1}$ is the right eigenvector associated with the eigenvalue 1 of $\boldsymbol{D}^{-1}\boldsymbol{A}$.

Each agent $i \in \mathcal{V}$ is associated with a continuously evolving opinion $x_i(t) \in \mathcal{X} := [-1, 1]$, representing their preference toward a certain behavior. Specifically, $x_i(t)$ is closer to -1 for pro-environmental attitudes and closer to 1 for non-environmental (or unsustainable) tendencies. The overall state of opinions is collected in the vector $\boldsymbol{x}(t) = (x_i(t))_{i\in\mathcal{V}} \in \mathcal{X}^N$.

The agents update their opinions by observing the behavior of their neighbors, which is captured by a continuously differentiable non-decreasing signal function $s : \mathcal{X} \to \mathcal{X}$. The collective perceived behaviors is given by the vector $\boldsymbol{s}(\boldsymbol{x})$ whose components are $s(x_i), i \in \mathcal{V}$.

The environment is modeled as a state $\tilde{e}(t) \in \mathbb{R}_{\geq 0}$, capturing the collective impact of the network's influence. The behavior of each agent i influences the environment through an increasing control function $u : \mathcal{X}^N \to [u_{\min}, u_{\max}]$, where $0 < u_{\min} < u_{\max}$. The environmental dynamics is:

$$\tau_e \dot{\tilde{e}} = -\gamma \tilde{e} + u(\boldsymbol{x}),$$

where $\gamma \in [0, 1]$ represents the environment's natural recovery rate (in the absence of human action) and $\tau_e > 0$ is a time constant reflecting the speed of the environmental dynamics.

Agents indirectly perceive the environmental state through a non-increasing function $r : \mathbb{R} \to \mathcal{X}$, which maps deviations from a threshold $\bar{e}$ into a response satisfying $r(\tilde{e} - \bar{e}) < 0$ if $\tilde{e} > \bar{e}$ and $r(\tilde{e} - \bar{e}) > 0$ if $\tilde{e} < \bar{e}$.

Assumption 1: The threshold $\bar{e}$ is such that $\bar{e} \in [u_{\min}/\gamma, u_{\max}/\gamma]$.

This assumption ensures that the environment is at the same scale as the threshold. This allows agents to influence the environment in a meaningful way [10].

For analytical convenience, we recenter the environmental variable by writing $\tilde{e} = \bar{e} + e$, where $e \in \mathbb{R}$ represents the deviation from the threshold. The environmental dynamics is then reformulated as:

$$\tau_e \dot{e} = -\gamma e + u(\boldsymbol{x}) - \gamma \bar{e}. \tag{1}$$

Accordingly, $r(e) > 0$ when $e < 0$ (i.e., when the environment is better than the threshold) and $r(e) < 0$ if $e > 0$.

The opinion of each agent evolves according to the dynamics:

$$\tau_x \dot{x}_i = -x_i + \beta r(e) + \frac{1-\beta}{d_i} \sum_{j=1}^{N} a_{ij} s(x_j), \tag{2}$$

where $\tau_x > 0$ is the time constant for the opinion dynamics and $\beta \in [0, 1]$ quantifies the trade-off between environmental and social influences. In particular, the parameter β plays a critical role: for $\beta < 0.5$, agents place more trust in the signals from their neighbors, whereas for $\beta > 0.5$ they give greater weight to the environmental signal. Similar OD models can be found in [10]–[12]. Such models can capture other behaviors than global synchronization like agreement, disagreement, clustering, and oscillations, making them encompass more realistic behaviors than the linear OD.

The coupled system (1) and (2) can be expressed in a matrix-vector form. Defining the system state as $\boldsymbol{y}(t) = (\boldsymbol{x}(t), e(t))$, the dynamics can be written in a compact form:

$$\dot{\boldsymbol{y}} = \boldsymbol{F}(\boldsymbol{y}) = \begin{bmatrix} \tau_x(-\boldsymbol{x} + \beta r(e)\mathbf{1} + (1-\beta)\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x})) \\ \tau_e(-\gamma e + u(\boldsymbol{x}) - \gamma \bar{e}) \end{bmatrix}. \tag{3}$$

III. Preliminary results

A. Forward Invariance of Synchronization manifold

We will now establish the forward invariance of the synchronization manifold

$$\mathcal{S} = \left\{ (\boldsymbol{x}, e) \in \mathcal{X}^N \times \mathbb{R} \mid \boldsymbol{x} = p\mathbf{1},\, p \in \mathcal{X} \right\}.$$

Proposition 1: Let $\mathcal{G}$ be a connected graph. Then, the synchronization manifold $\mathcal{S}$ is forward invariant for (3).

Proof: Let $(\boldsymbol{x}, e) \in \mathcal{S}$, i.e., $\boldsymbol{x} = p\mathbf{1}$ for some $p \in \mathcal{X}$. Then, for all $i, j \in \mathcal{V}$, one has

$$\begin{aligned} \dot{x}_i - \dot{x}_j &= \tau_x \left(\boldsymbol{e}_i - \boldsymbol{e}_j\right)^\top \left(\beta r(e)\mathbf{1} + (1-\beta)\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{x}\right) \\ &= \tau_x(1-\beta)\left(\boldsymbol{e}_i - \boldsymbol{e}_j\right)^\top \left(\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{s}(p\mathbf{1}) - p\mathbf{1}\right) \\ &= \tau_x(1-\beta)\left(s(p) - p\right)\left(\boldsymbol{e}_i - \boldsymbol{e}_j\right)^\top \mathbf{1} = 0. \end{aligned}$$

Since $\mathbf{1}$ is the eigenvector associated with the eigenvalue 1 of $\boldsymbol{D}^{-1}\boldsymbol{A}$, from Lemma 1. Thus, the synchronization manifold is forward invariant. ∎

The forward invariance of the synchronization manifold $\mathcal{S}$ ensures that once the system reaches the synchronization manifold, it remains there for all future times. Meaning that once all agents reach the same opinion, they will remain synchronized ever after. This property is crucial for the analysis of the system's behavior, as it allows us to focus on the dynamics over $\mathcal{S}$ and study the system's stability and bifurcations in a reduced space.

The attractiveness of the synchronization manifold $\mathcal{S}$ depends on the properties of the signal function s and graph topology. For example, as shown in [13, Proposition 3], for

global underestimation function, i.e., $x(s(x)-x) \leq 0$ for all $x \in [-1,1]$, with any connected graph the synchronization manifold is globally asymptotically stable. For more results on the attractiveness and attraction basin of $\mathcal{S}$, we refer to [13], where synchronization for the dynamics (2) have been studied without environment coupling (i.e, for $\beta = 0$).

B. Oddness of the dynamics

Assumption 2: The functions s, r, and u are smooth odd functions. Moreover, the control function u satisfies $u(\boldsymbol{x}) = -u(-\boldsymbol{x}) + 2\gamma\bar{e}$ for all $\boldsymbol{x} \in \mathcal{X}^N$.

Under Assumption 2, the dynamics (3) is odd. Indeed, for all $\boldsymbol{y} = (\boldsymbol{x}, e) \in \mathcal{X}^N \times \mathbb{R}$, one has:

$$\boldsymbol{F}(-\boldsymbol{y}) = \begin{bmatrix} \tau_x \boldsymbol{f}_1(-\boldsymbol{x}, -e) \\ \tau_e f_2(-\boldsymbol{x}, -e) \end{bmatrix} = \begin{bmatrix} -\tau_x \boldsymbol{f}_1(\boldsymbol{x}, e) \\ -\tau_e f_2(\boldsymbol{x}, e) \end{bmatrix} = -\boldsymbol{F}(\boldsymbol{y}).$$

This assumption is particularly useful. Indeed, in Section IV, we will analyze the dynamics (3) of the system over $\mathcal{S}$ using the Lyapunov-Schmidt reduction method [14]. This method requires, in general, a lot of computational effort since it involves a Taylor expansion of the third order of the Jacobian matrix. Under Assumption 2, the dynamics function exhibits an odd symmetry in the state variable, making the quadratic terms of this extension vanish and simplifying the analysis. Moreover, this assumption is also meaningful from a modeling perspective, as it aligns with the expectation that opposite behaviors occur around the neutral opinion and environment threshold.

IV. Analysis of synchronized agents dynamics

In this section, we assume that the states of the agents are identical, meaning that $\boldsymbol{x} \in \mathcal{S}$, i.e., there exists a $p \in \mathcal{X}$ such that $\boldsymbol{x} = p\mathbf{1}$. With a small abuse of notation, we will denote by $u(p)$ the function $u(p\mathbf{1})$. The dynamics with opinion in $\mathcal{S}$ is then given by:

$$\tau_x \dot{p} = -p + \beta r(e) + (1-\beta)s(p) \quad (4a)$$
$$\tau_e \dot{e} = -\gamma e + u(p) - \gamma\bar{e}. \quad (4b)$$

We call (4a) the Fully Synchronized Opinion (FSO) dynamics and denote $y = (p, e)$ its state.

Then, one can define the Fully Synchronized Opinion coupled with the Environment (FSOE) dynamics (4a)-(4b) through the function $F : \mathcal{X} \times \mathbb{R} \times [0,1]^2 \to \mathcal{X} \times \mathbb{R}$ as:

$$F(y, \beta, \gamma) = \begin{bmatrix} -p + \beta\, r(e) + (1-\beta)s(p) \\ -\gamma\, e + u(p) - \gamma\bar{e} \end{bmatrix}. \quad (5)$$

A. Equilibria

Let us define the instrumental function $g : \mathcal{X} \to \mathcal{X}$ as:

$$g(x) = \beta r\left(\frac{u(x)}{\gamma} - \bar{e}\right) + (1-\beta)s(x).$$

Proposition 2: Let $\boldsymbol{y}^*$ be an equilibrium of the FSOE dynamics. Then, $\boldsymbol{y}^* = (p^*, u(p^*)/\gamma)$ where p^* is a fixed point of g.

Proof: Let $\boldsymbol{y}^* = (p^*, u(x^*)/\gamma - \bar{e})$ be an equilibrium of the FSO dynamics. Then, one has the following from the environmental dynamics:

$$\dot{e} = 0 \Leftrightarrow e^* = u(p^*)/\gamma - \bar{e}.$$

Substituting this into the opinion dynamics yields:

$$\dot{p} = 0 \Leftrightarrow p^* = \beta r(e^*) + (1-\beta)s(p^*)$$
$$\Leftrightarrow p^* = \beta r\left(\frac{u(p^*)}{\gamma} - \bar{e}\right) + (1-\beta)s(p^*) \Leftrightarrow p^* = g(p^*).$$

■

The equilibrium points of the FSOE dynamics are then given by the fixed points of the instrumental function g. The function g may have multiple fixed points, leading to the existence of multiple equilibria of (4a)-(4b). In the following, we focus our analysis at $y^* = (0,0)$, that is always an equilibrium for any β and γ under Assumption 2.

B. Singular points

In this subsection, we analyze the conditions under which the Jacobian of the FSOE dynamics becomes singular. Identifying these singular points is essential as they mark the parameter values at which the linearization of the system loses full rank, thereby signaling potential bifurcations and qualitative changes in the system's behavior.

Our approach is as follows. First, we derive the explicit form of the Jacobian matrix of (5). In particular, establishing conditions under which the leading (i.e. maximal real part) eigenvalues of the Jacobian lie on the imaginary axispaves the way for rigorously proving the existence of pitchfork and Hopf bifurcations in later sections.

The Jacobian matrix of F is given by:

$$D_y F(y, \beta, \gamma) = \begin{bmatrix} \tau_x^{-1} & 0 \\ 0 & \tau_e^{-1} \end{bmatrix} \begin{bmatrix} (1-\beta)s'(p) - 1 & \beta r'(e) \\ u'(p) & -\gamma \end{bmatrix}.$$

The following proposition provides the conditions under which the Jacobian matrix of the FSOE dynamics has singular points. In the following, we will note $\tau = \tau_e/\tau_x$.

Proposition 3: The Jacobian matrix $D_y F(y, \beta, \gamma)$ has:

1) at least one eigenvalue equal to zero if and only if
$$\gamma = -\frac{u'(p)r'(e)\beta}{(1-\beta)s'(p) - 1}, \quad (6)$$
and $(1-\beta)s'(p) > \max(1, -u'(p)r'(e)\beta + 1)$.
2) two real eigenvalues equal to zero if and only if the conditions of 1) are satisfied and
$$\beta = 1 - \frac{1}{s'(p)} + \frac{-u'(p)r'(e) - \sqrt{\Delta_\beta}}{2\tau s'(p)^2},$$
with $\Delta_\beta = u'(p)r'(e)[u'(p)r'(e) - 4\tau s'(p)(s'(p)-1)]$.
3) two conjugated complex eigenvalues with to zero real part if and only if
$$\gamma = \tau((1-\beta)s'(p) - 1), \quad (7)$$

with $1 < (1-\beta)s'(p) < 1 + 1/\tau$ and $\beta \in (\beta_-, \min(\beta_+, 1))$, where

$$\beta_\pm = 1 - \frac{1}{s'(p)} + \frac{-u'(p)r'(e) \pm \sqrt{\Delta_\beta}}{2\tau s'(p)^2}.$$

Moreover, the eigenvalues are given by $\pm i\omega_0$ with $\omega_0 = \sqrt{\det(D_y F(y, \beta, \gamma))}$.

Proof: We prove each of the three statements in turn.

The Jacobian $D_y F(y, \beta, \gamma) := D_y F$ has a zero eigenvalue if and only if its determinant vanishes. A direct computation yields

$$\det(D_y F) = -\frac{\gamma\left((1-\beta)\, s'(p) - 1\right) + u'(p)r'(e)\beta}{\tau_x \tau_e}.$$

Thus, setting $\det(D_y F) = 0$ one obtains

$$\gamma = -\frac{\beta\, u'(p)\, r'(e)}{(1-\beta)s'(p) - 1}.$$

This relation is meaningful provided that $(1-\beta)s'(p) > \max(1, -u'(p)r'(e)\beta + 1)$ since $u'(p)r'(e) \leq 0$ and $\gamma \in [0, 1]$.

For the Jacobian to have a double zero eigenvalue, both the determinant and the trace must vanish. The trace is

$$\mathrm{Tr}(D_y F) = \frac{(1-\beta)s'(p) - 1}{\tau_x} - \frac{\gamma}{\tau_e}.$$

Then, the Jacobian has two eigenvalues equal to zero if and only if its trace and the determinant are equal to zero. This yields the following system of equations:

$$\gamma = -\frac{u'(p)r'(e)\beta}{(1-\beta)s'(p) - 1} \text{ and } \gamma = \tau((1-\beta)s'(p) - 1).$$

Then, one can retrieve the parameter β as the solution of the following equation:

$$\begin{aligned} &\tau((1-\beta)s'(p) - 1)^2 + u'(p)r'(e)\beta = 0 \\ \Leftrightarrow\ &\tau\beta^2 s'(p)^2 + \Big(u'(p)r'(e) - 2\tau s'(p)\left(s'(p) - 1\right)\Big)\beta \\ &\quad + \left(s'(p) - 1\right)^2 = 0. \end{aligned} \tag{8}$$

Denoting $\Delta_\beta = u'(p)r'(e)[u'(p)r'(e) - 4\tau s'(p)(s'(p) - 1)]$, one has that the equation has a real solution if and only if $u'(p)r'(e) \leq 4\tau s'(p)(s'(p) - 1)$. Under the assumption $u'(p)r'(e) \leq 0$ and $s'(p) > 1$ (implied by the condition in (1)), a real solution exists. The two solutions are given by

$$\beta_\pm = 1 - \frac{1}{s'(p)} + \frac{-u'(p)r'(e) \pm \sqrt{\Delta_\beta}}{2\tau s'(p)^2}.$$

A further inspection shows that only β_- satisfies the additional requirement $(1-\beta)s'(p) > 1$.

Finally, the Jacobian has two complex conjugate eigenvalues with zero real part if and only if (i) the trace vanishes and (ii) the discriminant of the characteristic polynomial is negative. The characteristic polynomial is

$$P(X) = X^2 - \mathrm{Tr}(D_y F)\, X + \det(D_y F)\,.$$

with discriminant

$$\Delta_P = \mathrm{Tr}(D_y F)^2 - 4\det(D_y F)\,. \tag{9}$$

Setting $\mathrm{Tr}(D_y F) = 0$ again yields

$$\gamma = \tau((1-\beta)s'(p) - 1),$$

Moreover, to ensure that the eigenvalues are non-real, we require $\Delta_P < 0$, which is equivalent to $\det(D_y F) > 0$. This inequality leads to

$$\gamma\left((1-\beta)\, s'(p) - 1\right) + u'(p)r'(e)\beta < 0.$$

In light of our previous derivations, this condition is satisfied for $\beta \in (\beta_-, \min\{\beta_+, 1\})$, with the additional constraint that $1 < (1-\beta)s'(p) < 1 + 1/\tau$. Moreover, the eigenvalues are given by $\pm i\omega_0$, where $\omega_0 = \sqrt{\det(D_y F(y, \beta, \gamma))}$. ■

This proposition provides a complete characterization of the singular point of the Jacobian matrix of the FSOE dynamics (4a).

Remark 1: It is noteworthy that the condition $s'(p) > 1$ is necessary for the Jacobian of the FSOE dynamics at $y = (p, e)$ to become singular. In other words, at the opinion state p, the signal function s must act as an amplifier of the agents' opinions for it to be singular.

C. Bifurcation analysis

The following result provides conditions for a pitchfork bifurcation at the equilibrium $(p, e) = (0, 0)$ of (4a)-(4b).

Theorem 1: Suppose Assumption 2 holds and that β^* and $\gamma^* = \gamma(\beta^*)$ satisfy (6) at $(p, e) = (0, 0)$. Then, the Jacobian of the FSOE dynamics (4a)-(4b) at the equilibrium $y^* = (0, 0)$ has a zero eigenvalue associated with the critical eingenvector

$$v = \begin{bmatrix} 1 & u'(0)/\gamma^* \end{bmatrix}^\top.$$

Moreover, if

$$c = (1-\beta^*)s'''(0) + \beta^* r'''(0)\frac{u'(0)^3}{{\gamma^*}^3} + \frac{\beta^* r'(0)}{\gamma^*}u'''(0)$$

is nonzero, then a pitchfork bifurcation occurs at y^* as β passes through β^*. In particular, the bifurcating branches emerge along the subspace generated by v.

Proof: From proposition 3, when β^* verifies (6), we know that the Jacobian of the FSOE dynamics (4a)-(4b) at the equilibrium $y^* = (0, 0)$ is singular. Inspired by [15], we perform the Lyapunov-Schmidt reduction of the equilibrium equation $F(y, \beta^*, \gamma^*) = 0$. At (y^*, β^*, γ^*), the linearization $D_y F$ is singular and has a one-dimensional kernel. Let v and w be the corresponding right and left eigenvectors associated with the zero eigenvalues. One has

$$v = \begin{bmatrix} 1 & u'(0)/\gamma^* \end{bmatrix}^\top \quad \text{and} \quad w = \begin{bmatrix} 1 & \beta^* r'(0)/\gamma^* \end{bmatrix}.$$

Then, $P = ww^\top / \|w\|$ is the projection onto the kernel of $D_y F(y^*, \beta^*, \gamma^*)$ and $Q = I - P$ is the projection onto its range. One can write a general nearby solution as $y = zv + \xi w$, where $z \in \mathbb{R}$ the coordinate along the kernel and $\xi \in \mathbb{R}$ the coordinate along the range. Let $\mu = \beta - \beta^*$ denote the unfolding parameter. The equilibrium equation $F(y, \beta^*, \gamma^*) = 0$ is then equivalent to the system

$PF(y, \beta^* + \mu, \gamma^*) = 0$ and $QF(zv + \xi w, \beta^* + \mu, \gamma^*) = 0$.

By construction, one has $QF(y^*, \beta^*, \gamma^*) = 0$. Since $QD_yF(y^*, \beta^*, \gamma^*)$ is invertible, the implicit function theorem ensures that there exists a unique function $h(z, \mu)$ such that $QF(zv + h(z, \mu)w, \beta^* + \mu, \gamma^*) = 0$ and $h(0, 0) = 0$. Then, a nearby solution to the equilibrium equation is given by $y = zv + h(z, \mu)$.

Substituting this ansatz into $F(y, \beta^* + \mu, \gamma^*) = 0$ and projecting onto the kernel (by multiplying on the left by $w^\top$) yields the reduced bifurcation equation $\Phi(z, \mu) = 0$.

Since the functions s, r, and u are odd, their Taylor expansions about the origin contain only odd-order terms. This symmetry ensures that no quadratic term appears in the expansion of Φ in z. Thus, expanding in powers of z and μ we obtain

$$\Phi(z, \mu) = a\mu z + \frac{1}{6}cz^3 + \mathcal{O}(z^5, \mu z^3, \mu^2 z) = 0,$$

with some constant

$$a = w^\top \partial_\beta F(y^*, \beta^*, \gamma^*) = w^\top \begin{bmatrix} \frac{r(0)-s(0)}{\tau_x} & 0 \end{bmatrix}^\top = 0,$$

since $r(0) = s(0) = 0$. The third-order terms are given by:

$$D_y^3 F(y^*, \beta^*, \gamma^*)(h, k, l) = \begin{bmatrix} \tau_x^{-1} & 0 \\ 0 & \tau_e^{-1} \end{bmatrix} \begin{bmatrix} (1-\beta^*)s'''(0)h_1k_1l_1 \\ + \beta^* r'''(0)h_2k_2l_2 \\ u'''(0)h_1k_1l_1 \end{bmatrix}$$

for $y = (p, e)$, $h = (h_1, h_2)$, $k = (k_1, k_2)$ and $l = (l_1, l_2)$.

Then, the cubic coefficient c is given by

$$\begin{aligned} c &= w^\top \left[D_y^3 F(y^*, \beta^*, \gamma^*)(v, v, v) \right] \\ &= \frac{(1-\beta^*)s'''(0)}{\tau_x} + \frac{\beta^* r'''(0)u'(0)^3}{\tau_x {\gamma^*}^3} + \frac{\beta^* r'(0)u'''(0)}{\tau_e \gamma^*}. \end{aligned}$$

Consequently, if $c \neq 0$ a pitchfork bifurcation occurs at y^* as $\beta = \beta^*$. Precisely, two nontrivial symmetric solutions bifurcate from $z = 0$, along the v direction, when the sign of μ changes. ∎

From an environmental view-point, this pitchfork bifurcation captures the idea that small changes in the trust parameter β can trigger sudden transitions between collective opinion states. Close to a pitchfork bifurcation, dynamics are bistable: the state can converge to either of two stable equilibria, depending on perturbations or initial conditions. In the environmental setting, either the collective opinion strongly favors pro-environmental actions, resulting in a well-preserved environment, or the prevailing sentiment opposes environmental efforts, leading to environmental degradation.

Remark 2: Although the pitchfork of Theorem 1 happens at an unstable trivial equilibrium, it is this bifurcation that gives rise to bistability in the FSOE dynamics.

In addition to the pitchfork bifurcation, the trivial equilibrium also exhibits a Hopf bifurcation for a different range of parameter β.

Theorem 2: Suppose Assumption 2 holds and let β^* and $\gamma^* = \gamma(\beta^*)$ satisfy (7) with equilibrium $y^* = (0, 0)$. Then, a Hopf bifurcation yielding a family of periodic orbits occurs at y^* as $\beta = \beta^*$. Moreover, the bifurcating limit cycle is unique and stable for $\beta < \beta^*$ if the coefficient h_{21} defined in (11) satisfies $\mathrm{Re}(h_{21}) \neq 0$.

Proof: By [14, Theorem 3.4.2], the existence of periodic orbits follows if the system (4a)-(4b) satisfies two conditions at the equilibrium $y^* = (0, 0)$: the eigenvalues of the Jacobian matrix $D_yF(y^*, \beta^*, \gamma^*)$ are purely imaginary, and the derivative of the real part of the eigenvalues with respect to β is nonzero.

First, by Proposition 3, the eigenvalues of the Jacobian matrix at $y^* = (0, 0)$ are purely imaginary, satisfying the first condition. Moreover, in a neighborhood of (y^*, β^*, γ^*), the characteristic equation of the Jacobian D_yF has discriminant (9) with $\Delta_P < 0$, giving eigenvalues

$$\lambda = \frac{\mathrm{Tr}(D_yF(y, \beta, \gamma^*)) + i\sqrt{-\Delta_P(y, \beta, \gamma^*)}}{2},$$

with Δ_P given by (9).

The second condition is also satisfied since the real part of these eigenvalues is $\mathrm{Re}(\lambda) = \mathrm{Tr}(D_yF(y, \beta, \gamma^*))/2$. Differentiating with respect to β, we obtain $\partial_\beta \mathrm{Re}(\lambda) = -s'(0)/\tau_x$, which is nonzero since $s'(0) > 1$. This proves the existence of the periodic orbits.

The stability and uniqueness of the limit cycle results from [16, Theorem 3.3]. First, let us provide the normal form of the bifurcation. At (y^*, β^*, γ^*), the linearization D_yF is singular and has a two-dimensional center subspace associated with the purely imaginary eigenvalues $\pm i\omega_0$. Let q and $\bar{q}$ be the right eigenvectors and p and $\bar{p}$ be the left eigenvectors associated with the eigenvalues $i\omega_0$ and $-i\omega_0$, respectively. They are given by

$$q = \begin{bmatrix} \tau_x^{-1}\beta^* r'(0) \\ i\omega_0 - a \end{bmatrix} \quad \text{and} \quad p = \begin{bmatrix} \tau_e^{-1}u'(0) & i\omega_0 - a \end{bmatrix},$$

where $a = \tau_x^{-1}((1-\beta^*)s'(0) - 1) = \tau_x^{-1}\gamma^*$.

By [16, Lemma 3.3], for β sufficiently close to β^* and setting $\mu = \beta - \beta^*$, the FSOE dynamics (4a)-(4b) can be transformed via a complex variable z into

$$\dot{z} = \lambda z + h(z, \bar{z}, \mu), \tag{10}$$

with $h(z, \bar{z}, \mu) = \mathcal{O}(|z|^2)$ a smooth function of z and $\bar{z}$.

Expanding h in powers of z and $\bar{z}$,

$$h(z, \bar{z}, \mu) = \sum_{k+l \geq 2} \frac{1}{k!l!} \frac{\partial^{k+l}}{\partial z^k \, \partial \bar{z}^l} \left\langle p, F\big(zq + \bar{z}\bar{q}, \mu\big) \right\rangle_{|z=0} z^k \bar{z}^l.$$

By [16, Lemma 3.6], equation (10) can be rewritten as $\dot{z} = \lambda z + c_1 z^2 \bar{z} + \mathcal{O}(|z|^4)$, where c_1 is the first Lyapunov exponent given by

$$c_1 = \frac{h_{20}h_{11}(2\lambda + \bar{\lambda})}{2\,|\lambda|} + \frac{|h_{11}|^2}{2} + \frac{|h_{02}|^2}{2(2\lambda - \bar{\lambda})} + \frac{h_{21}}{2}.$$

Since s, r and u are odd functions, one has that $h_{kl} = 0$ for $k + l$ even, leading to $c_1 = h_{21}/2$. The coefficient h_{21} is

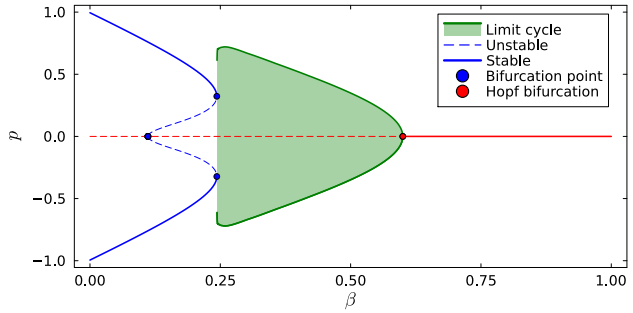


Fig. 1: Bifurcation diagram for the FSO dynamics (4a) with $\gamma = 0.2$. Solid lines denote stable equilibria, dashed lines denote unstable equilibria, dots mark bifurcation points, and the green region shows the amplitude of the stable limit cycle emerging from the Hopf bifurcation.

given by

$$h_{21} = \bar{p}^\top D_y^3 F(y^*, \beta^*, \gamma^*)(q, q, \bar{q}) = u'(0)\left[\frac{(1-\beta^*)s'''(0)\left(\beta^* r'(0)\right)^3}{\tau_x^3} + \beta^* r'''(0)(i\omega_0 + a)(i\omega_0 - a)^2\right] - \frac{u'''(0)(i\omega_0 + a)(\beta^* r'(0))^3}{\tau_x^2}, \tag{11}$$

with $a = \tau_x^{-1}\gamma^*$ and $\omega_0 = \sqrt{\det(D_y F(y^*, \beta^*, \gamma^*))}$. If $\mathrm{Re}(h_{21}) \neq 0$ one has $\mathrm{Re}(c_1) \neq 0$ and from [16, Theorem 3.3] there is a unique stable limit cycle that bifurcates from the equilibrium y^* via a Hopf bifurcation for $\beta < \beta^*$. ■

In summary, our bifurcation analysis reveals two distinct types of qualitative transitions in the coupled opinion-environment dynamics. The pitchfork bifurcation indicates a sudden, symmetry-breaking shift in collective behavior, whereas the Hopf bifurcation signals the emergence of oscillatory dynamics, which may model the recurrent cycles of environmental collapse and recovery observed in prey-predatory systems.

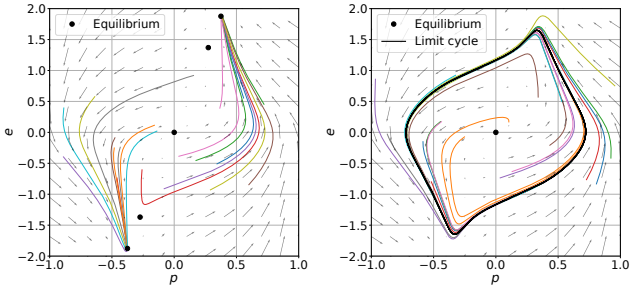
V. NUMERICAL SIMULATIONS

We illustrate the results of Theorem 1 and 2 with numerical simulations. We consider the following functions $s(x) = \tanh(3x)$, $r(x) = \tanh(-3x)$ and $u(x) = x - \gamma\bar{e}$, where $\bar{e} = 0.5$ is the environmental threshold. We set $\tau_x = \tau_e = 1$ and $\gamma = 0.2$. We compute the bifurcation diagram for the FSO dynamics (4a) with respect to the parameter β in Figure 1. We observe a pitchfork bifurcation around $\beta = 0.24$ and a Hopf bifurcation around $\beta = 0.60$ in Figures 2 and 3, respectively.

Remark 3: As illustrated in Figure 1, the system also exhibits saddle-node bifurcation involving the equilibria that emerge from the pitchfork bifurcation of Theorem 1. Due to lack of space, we omit a detailed analysis of these saddle-node bifurcations; such an analysis could be carried out using techniques analogous to those in Theorem 1.

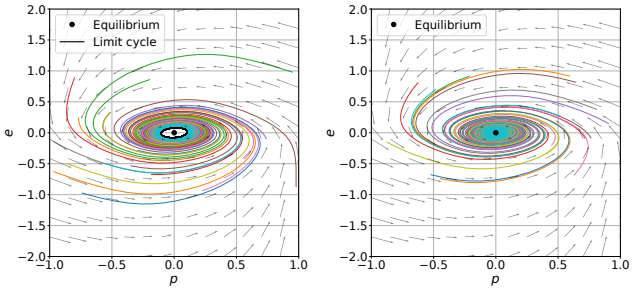
VI. CONCLUSIONS

In this paper, we introduce and analyze a continuous-time opinion-environment model as an extension of [10],



(a) $\beta = 0.24$ (b) $\beta = 0.25$

Fig. 2: Phase portraits for two value of β and $\gamma = 0.2$. The system passes from 5 equilibria for $\beta = 0.24$ to 1 equilibrium with a stable limit cycle for $\beta = 0.25$.



(a) $\beta = 0.59$ (b) $\beta = 0.61$

Fig. 3: Phase portraits for two value of β and $\gamma = 0.2$. The stable limit cycle for $\beta = 0.59$ collapses into an equilibrium at $\beta = 0.61$ illustrating the Hopf bifurcation of Figure 1.

capturing the interplay between social interactions and environmental feedback. We establish positive invariance, explore singularities of the FSOE dynamics, and demonstrate pitchfork and Hopf bifurcations using a rigorous mathematical framework.

REFERENCES

[1] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of Artifical Societies and Social Simulation (JASSS) vol*, vol. 5, no. 3, 2002.

[2] N. E. Friedkin and E. C. Johnsen., "Social influence and opinions." *Journal of Mathematical Sociology.*, vol. 15, pp. 193–206, 1990.

[3] I.-C. Morărescu and A. Girard, "Opinion dynamics with decaying confidence," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1862 – 1873, 2011.

[4] A. Bizyaeva, A. Franci, and N. E. Leonard, "Nonlinear Opinion Dynamics With Tunable Sensitivity," *IEEE Transactions on Automatic Control*, vol. 68, no. 3, pp. 1415–1430, Mar. 2023.

[5] S. Martin, I.-C. Morărescu, and D. Nešić, "Consensus and influence power approximation in time-varying and directed networks subject to perturbations," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 11, pp. 3485 – 3501, 2019.

[6] C. Bernardo, L. Wang *et al.*, "Quantifying leadership in climate negotiations: A social power game," *PNAS*, vol. 2, no. 11, p. 365, 2023.

[7] J. S. Weitz, C. Eksin *et al.*, "An oscillating tragedy of the commons in replicator dynamics with game-environment feedback," *PNAS*, vol. 113, no. 47, pp. 7518–7525, 2016.

[8] A. R. Tilman, J. B. Plotkin, and E. Akçay, "Evolutionary games with environmental feedbacks," *Nature Communications*, vol. 11, no. 1, p. 915, Feb. 2020.

[9] K. Frieswijk, L. Zino *et al.*, "Modeling the Co-evolution of Climate Impact and Population Behavior: A Mean-Field Analysis," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 7381–7386, Jan. 2023.

[10] A. Couthures, V. Satheeskumar Varma *et al.*, "Analysis of an opinion dynamics model coupled with an external environmental dynamics," *Chaos, Solitons & Fractals*, vol. 189, p. 115719, Dec. 2024.
[11] A. Bizyaeva, A. Franci, and N. E. Leonard, "Multi-Topic Belief Formation Through Bifurcations Over Signed Social Networks," *IEEE Transactions on Automatic Control*, pp. 1–16, 2025.
[12] R. Gray, A. Franci *et al.*, "Multiagent Decision-Making Dynamics Inspired by Honeybees," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 2, pp. 793–806, Jun. 2018.
[13] A. Couthures, V. S. Varma *et al.*, "Global synchronization of multi-agent systems with nonlinear interactions," (submitted) 2025.
[14] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, ser. Applied Mathematical Sciences. Springer, 1983, vol. 42.
[15] M. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, ser. Applied Mathematical Sciences, J. E. Marsden, L. Sirovich, and F. John, Eds. Springer New York, 1985, vol. 51.
[16] Y. A. Kuznetsov, "Elements of Applied Bifurcation Theory," *Second Edition*, 1998.

Appendix B

Paper of the first stage: Formalizing Strategic Conflict

International Climate Policy as a Potential Game: Strategic Incentives and Path-Dependent Dynamics

Anthony Couthures[a], Bouchra Mroué[a], Samson Lasaulce[a], Irinel-Constantin Morărescu[a,b]

[a] *Université de Lorraine, CNRS, CRAN, Nancy, F-54000, France*
[b] *Automation Department, Technical University of Cluj-Napoca, Cluj-Napoca, Romania*

Abstract

The strategic challenge of climate policy lies in designing international agreements that handle the tensions inherent in managing a global public good, where the uneven distribution of costs and benefits creates incentives for free-riding. While the game-theoretic literature on International Environmental Agreements (IEAs) provides analytical clarity on coalition stability, Integrated Assessment Models (IAMs) capture complex geophysical dynamics but often abstract from strategic incentives. This paper bridges these approaches by embedding a nonlinear, path-dependent geophysical module directly into the utility structure of a dynamic, open-loop game of international climate policy. We show that, despite the complexity introduced by the climate module, the resulting interaction forms a state-transitive weighted potential game, guaranteeing the existence of a pure-strategy Nash Equilibrium. Using this framework, we generate three key insights. First, the classic "paradox of cooperation" is endogenous to the perception of catastrophic climate economic damage: higher perceived damage can strengthen incentives for cooperation. Second, truncating analysis at 2100 can dramatically underestimate the true strategic incentives for both cooperative and unilateral action, highlighting the importance of long-term foresight. Third, adaptive institutional mechanisms, akin to the ratchet structure of the Paris Agreement, are superior to static long-term commitments in achieving robust and effective climate outcomes. Our results demonstrate that integrating realistic geophysical risk into strategic models provides a tool for understanding and designing climate agreements, offering both theoretical and policy-relevant insights for addressing the global climate challenge.

Keywords: International Environmental Agreements, Noncooperative Games, Potential Games, Catastrophic Climate Damages, Strategic Foresight

1. Introduction

The central challenge of climate policy in the era of the Paris Agreement is designing international agreements that can navigate the strategic tensions emerging in managing a global public good. Mitigating climate change requires a coordinated, global effort, yet the costs and benefits of action are unevenly distributed, creating important incentives for nations to free-ride on the abatement efforts of others [1, 2]. This core dilemma has animated decades of research and negotiation, seeking to understand the conditions under which international cooperation can emerge and be sustained [3].

The foundational game-theoretic literature on International Environmental Agreements (IEAs) offers a central, if pessimistic, baseline for this problem. Seminal work by [4] and [5] famously demonstrates the "paradox of cooperation": when the potential gains from cooperation are large, the incentive to free-ride is also at its strongest, leading to stable coalitions that are too small to be effective [6]. This result has been rigorously confirmed even decades later [7]. Because IEAs must be self-enforcing in the absence of an international authority, understanding the stability

⋆This work was supported by the project DECIDE funded under the PNRR I8 scheme by the Romanian Ministry of Research. The authors declare that they have no conflicts of interest related to the research described in this paper. The authors have no financial interest to disclose.

Email addresses: anthony.couthures@univ-lorrraine.fr (Anthony Couthures), bouchra.mroue@gmail.com (Bouchra Mroué), samson.lasaulce@univ-lorraine.fr (Samson Lasaulce), constantin.morarescu@univ-lorraine.fr (Irinel-Constantin Morărescu)

conditions of coalitions is capital [8]. The comprehensive body of work in this area suggests that achieving broad, meaningful participation is a complex strategic challenge [9, 10].

However, the study of climate policy has largely evolved along two large but often parallel tracks, creating a disciplinary fracture. While the game-theoretic literature prioritizes strategic abstraction, the field of Integrated Assessment Models (IAMs) offers a comprehensive, simulation-based approach. Pioneered by the DICE model of [11] for the cooperative framework and the RICE model [12], IAMs couple a classical economic growth framework with simplified representations of the climate system, a methodology detailed in works like [2] and further developed over decades [13]. These models are important for the scenario analysis that constructs major policy assessments, such as those by the Intergovernmental Panel on Climate Change (IPCC), which relies on a panel of IAMs to explore mitigation pathways [14, 15]. Recognizing this fracture, some research has tried to bridge the gap by embedding coalition formation analysis within IAMs, though typically through numerical simulation rather than formal analytical proofs [16, 17].

This disciplinary fracture is not merely a matter of epistemology; it has profound implications. The analytical clarity of IEA models is often purchased at the price of abstracting away from crucial geophysical dynamics. Conversely, while "black box" IAMs capture more physical complexity, their reliance on stylized, often quadratic, damage functions has been a subject of intense critique [18, 19, 20, 21]. Moreover, until recently, the environmental model of IAMs were not aligned with the physical reality of the climate system, leading to a significant disconnect between the economic and physical models [22, 23, 24]. A more fundamental concern, articulated in the seminal work of [25], is that these models are ill-suited to capture catastrophic economic damages. The author argues that deep structural uncertainty about the climate-economy relationship implies a "fat tail" of low-probability, high-damage outcomes that could dominate any rational cost-benefit analysis [26, 27]. This theoretical concern is supported by scientific evidence identifying multiple potential "tipping elements" in the Earth's climate system [28, 29], whose inclusion has been shown to substantially increase the economic damages of climate change [30]. This raises a critical, unanswered question: how do strategic incentives for cooperation change when the possibility of such nonlinear, catastrophic damages is embedded within a formal, analytically tractable game-theoretic framework?

Efforts to move beyond the static IEA framework have largely followed two paths. One approach incorporates dynamics through the lens of differential games, typically analyzing Markov Perfect Equilibria in the context of stock pollutants [31, 32, 33]. A second, more recent line of research focuses on the role of direct political and social influence. For instance, [34] models cooperation as an "imitation game," demonstrating that exogenous peer pressure and network topology can be main drivers of cooperative behavior, helping to overcome the free-rider problem.

This paper proposes a novel methodological path that bridges the gap between the IEA and IAM approaches by coupling the analytical rigor of game theory with the geophysical realism of IAMs. We focus on the endogenous strategic incentives created by the climate system itself. We develop and analyze a dynamic, open-loop game of international climate policy that embeds a path-dependent geophysical module coupled with nonlinear damage dynamics directly into players' utility functions. This module is built using the structure of IPCC-vetted Simple Climate Models (SCM) [35, 36, 37]. Our central theoretical contribution is to prove that, despite the introduction of nonlinear and path-dependent dynamics, the resulting game possesses the structure of a *state-transitive weighted potential game* [38, 39, 40]. This structure, distinct from both the imitation-game and standard differential-game approaches, is the key that unlocks our analysis, rendering an otherwise intractable problem analytically tractable.

Leveraging the analytical tractability of this potential game framework, we generate three main findings. First, we demonstrate that the "paradox of cooperation" is not a universal constant but is *endogenous* to the shared perception of physical risk, providing a direct strategic consequence to the concerns raised by [26]. Second, we reveal the critical role of strategic foresight and the potential for a "tyranny of the terminal horizon", showing that the common practice of truncating analysis at 2100 may severely underestimate the true economic incentives for both cooperation and unilateral action. Finally, we show that institutional frameworks similar to the "ratchet mechanism" of the Paris Agreement, that allow for adaptive, iterative re-commitments, are game-theoretically superior to both *myopic* policies (which ignore the future) and static *long-term commitments* (which suffer from end-of-horizon effects).

The remainder of this paper is structured as follows. Section 2 develops our climate-economic framework, linking country-level benefit and damage functions to a nonlinear geophysical model, and formally defines the interaction as a dynamic, open-loop game. Section 3 presents our central theoretical result: we prove that this complex game possesses the structure of a state-transitive weighted potential game, which directly guarantees the existence of a pure-strategy Nash Equilibrium (NE). Section 4 then investigates the properties of this equilibrium, providing a full characterization

via the Karush-Kuhn-Tucker (KKT) conditions and establishing sufficient conditions for its uniqueness.

Having characterized the non-cooperative outcome, Section 5 analyzes the core problem of cooperation. We formally demonstrate the inefficiency of the Nash Equilibrium relative to the Social Optimum (SO) and prove the inherent instability of the fully cooperative Grand Coalition, thereby deriving the foundational "paradox of cooperation" within our dynamic setting. Finally, Section 6 presents a detailed numerical analysis that calibrates the model to real-world data, quantifies our theoretical findings, and explores the strategic implications of perceived economic damage from climate change, foresight, and alternative policy frameworks. Section 7 concludes with a discussion of policy implications and directions for future research.

2. A Game of Emissions and Climate Dynamics

In this section, we construct a mathematical model (namely, a game model) to analyze the interactions between a set of countries that produce emissions. Constructing such a game amounts to modeling the behavior of the different countries involved. The proposed construction aims to find a good tradeoff between tractability (to allow both the mathematical and numerical analyses to be conducted) and realism (by integrating key components that drive decision-making). More specifically, we develop a game-theoretic model of international climate policy that links country-level economic decisions to a dynamic, nonlinear geophysical model of the global climate. The model is structured around three core components: the economic environment defining players' costs and benefits, the geophysical module converting the total CO_2 emissions into temperature change, and the formal game structure defining the strategic interaction.

2.1. The Economic Environment

We consider a global economy composed of a set of N countries, indexed by $n \in \mathcal{N} \coloneqq \{1, \dots, N\}$. The game takes place over a discrete-time horizon of T periods, indexed by $t \in \mathcal{T} \coloneqq \{1, \dots, T\}$. For example, when modeling climate change in five-year increments over the 2020–2100 horizon, we set $T = 16$. In each period, country n chooses an action a_{nt}, representing its level of Greenhouse Gas (GHG) emissions, in CO_2 equivalent. Each country's choice is constrained to a compact and convex action set, $a_{nt} \in \mathcal{A}_n \coloneqq [e_n^{\min}, e_n^{\max}]$, $e_n^{\min} \geq 0$. Let $\boldsymbol{a}_t = (a_{1t}, \dots, a_{Nt})$ denote the profile of actions/emissions in period t.

Benefits from Emissions. Emissions are a byproduct of productive economic activities. We model this relationship using a country-specific benefit function, $B_n(a_{nt})$. This function serves as a reduced-form representation of the gross economic welfare (e.g., GDP) generated from emitting a_{nt}. We impose standard regularity conditions:

Assumption 1 (Benefit Function)**.** *For all $n \in \mathcal{N}$, the benefit function $B_n : \mathcal{A}_n \to \mathbb{R}$ is twice-continuously differentiable, strictly increasing ($B_n' > 0$), and strictly concave ($B_n'' < 0$).*

Assuming the benefit function to be increasing with the emission level is reasonable and has been motivated in various studies (see e.g., [41, 42] for pioneering papers and [43, 44] for more recent works). Strict concavity reflects diminishing marginal economic returns to emissive activities, a standard feature in IAMs where abatement becomes progressively more expensive. Note that some results provided in this paper, such as the existence of a pure Nash equilibrium, do not require these assumptions.

Damages from Climate Change. Global GHG emissions contribute to a rise in global mean temperature [45, 46], which in turn causes economic damages. This notion has been widely accepted by the economic literature (e.g., see [47, 12, 27, 48]) but has started to be quantified after concrete climate change catastrophes (e.g., see [49, 50, 51]). Climate change is, except for some specific countries[1], a pure public bad; all countries are affected by the global temperature, regardless of their individual contribution. We model this through a global damage function, D, which is driven by the global mean atmospheric temperature anomaly, θ^{at}. To account for the fact that countries exhibit heterogeneous vulnerabilities to such a global temperature change, we scale these impacts using a set of strictly positive weights, $w_n > 0$. The total damage incurred by country n is thus $w_n D(\theta^{\text{at}})$:

[1]For instance, precipitation levels in the Sahara Desert are projected to increase, while the rise in average temperatures in Siberia may be perceived as a local climatic advantage: `https://interactive-atlas.ipcc.ch/`.

Assumption 2 (Damage Function)**.** *The global damage function $D : \mathbb{R} \to \mathbb{R}$ is twice-continuously differentiable, strictly increasing ($D' > 0$), and strictly convex ($D'' > 0$).*

Strict convexity of the damage function captures the widely accepted assumption, that is supported by empirical and process-based studies [52, 53], that economic damages from climate change accelerate as atmospheric temperature rises (due to well-known positive feedbacks like ice-albedo [54], permafrost [55], or Amazon deforestation [56]).

2.2. Geophysical Dynamics: From Emissions to Temperature

A key feature of our model is the explicit dynamic and nonlinear relation between aggregate GHG emissions choices $E_t = \sum_{n=1}^{N} a_{nt}$ and the resulting change in global mean atmospheric temperature θ^{at}_{t+1}. This process is governed by a standard Simple Climate Model (SCM)[2], consistent with the structure of models used by the IPCC [37, 35].

Let the state of the geophysical system at the beginning of period t be described by a state vector $\boldsymbol{x}_t$. The details about the system state can be found in Appendix A.1; the state is a vector made of various CO_2 concentrations (for the atmosphere and ocean typically) and temperatures (also for the atmosphere and ocean). The model translates emissions into temperature through a two-step process, which can be summarized by the following reduced-form equations.

First, the atmospheric CO_2 quantity at the end of period $t + 1$, C^{at}_{t+1}, is a linear function of the aggregate emissions within the period and the inherited state of the carbon cycle:

$$C^{\text{at}}_{t+1} = \psi_C(\boldsymbol{x}_t) + \beta_{\text{C}} E_t. \tag{1}$$

Here, $\psi_C(\boldsymbol{x}_t)$ represents the evolution of atmospheric carbon based on the initial state $\boldsymbol{x}_t$, while the parameter $\beta_{\text{C}} > 0$ captures the marginal impact of current emissions.

Second, the global mean atmospheric temperature at the end of period $t + 1$, θ^{at}_{t+1}, is a function of this new carbon state and the preceding state. This relationship is nonlinear by nature due to the physics of radiative forcing:

$$\theta^{\text{at}}_{t+1} = \psi_\theta(\boldsymbol{x}_t) + \beta_\theta \ln(C^{\text{at}}_{t+1}). \tag{2}$$

The term $\psi_\theta(\boldsymbol{x}_t)$ captures the system's thermal inertia and the influence of exogenous factors, while the parameter $\beta_\theta > 0$ represents the temperature's sensitivity to atmospheric carbon. The logarithmic form is a direct consequence of the physics of radiative forcing and is a critical feature of our analysis.

A detailed presentation of the SCM's state–space formulation can be found in Appendix A.1.

Remark 1 (Uniformly Mixed Pollutant)**.** *The structure of the geophysical model implies that the marginal effect of any country's emissions on global temperature is identical for all countries, i.e., $\partial\theta^{\text{at}}_{t+1}/\partial a_{nt} = \partial\theta^{\text{at}}_{t+1}/\partial a_{kt}$ for all $n, k \in \mathcal{N}$. This formalizes the public bad nature of climate change and is a key driver of the strategic incentives in the game.*

A full derivation of the functions ψ_C, ψ_θ, and the parameters $\beta_{\text{C}}, \beta_\theta$ from the underlying linear state-space models of the carbon and temperature cycles is provided in Appendix A.2.

2.3. Utilities, Strategies, and Equilibrium

We now combine the economic and geophysical components to formally define the game of interest. The instantaneous utility for country n in period t is the difference between its benefits from emissions and the economic damages caused by the resulting temperature increase:

$$u_{nt}(\boldsymbol{a}_t; \boldsymbol{x}_t) = B_n(a_{nt}) - w_n D\big(\theta^{\text{at}}_{t+1}(\boldsymbol{a}_t; \boldsymbol{x}_t)\big), \tag{3}$$

where the notation θ^{at}_{t+1} makes explicit the dependence of the end-of-period temperature on the actions taken within the period, $\boldsymbol{a}_t$, and the state at the beginning of the period, $\boldsymbol{x}_t$.

[2] Simple in the computational complexity [24].

A pure strategy for player n is a complete action plan of emissions over the entire time horizon, denoted by the vector $\boldsymbol{a}_n = (a_{n1}, \ldots, a_{nT}) \in \mathcal{A}_n^T := \prod_{t=1}^T \mathcal{A}_n$ (note that this vector implicitly depend on the horizon T but it is omitted for notation simplification since T is not varying). The set of all feasible strategies for all players is given by n is $\mathcal{A}^T := \prod_{n=1}^N \mathcal{A}_n^T$. The total utility for player n for a given strategy profile $\boldsymbol{a} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N) \in \mathcal{A}^T$ is the sum of its instantaneous utilities:

$$U_n(\boldsymbol{a}; \boldsymbol{x}_1) = \sum_{t=1}^{T} u_{nt}(\boldsymbol{a}_t; \boldsymbol{x}_t), \tag{4}$$

where the state $\boldsymbol{x}_t$ for $t > 1$ is endogenously determined by the sequence of actions $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{t-1})$ evolving from the initial state $\boldsymbol{x}_1$. Note that the utility function depends on the horizon parameter T, but is omitted for notation simplification for the same reason as the action profiles.

We analyze this interaction as an open-loop game where all players simultaneously choose their entire strategy vector $\boldsymbol{a}_n$ at the beginning of the game.

Definition 1 (Open-Loop Climate Policy Game)**.** *The game Γ_T is defined by the tuple $\Gamma_T := (\mathcal{N}, \{\mathcal{A}_n^T\}_{n\in\mathcal{N}}, \{U_n\}_{n\in\mathcal{N}})$.*

The game has a distinctive structure. While the physical environment is dynamic, with a state that evolves in a path-dependent manner, the strategic interaction is modeled as a one-shot, simultaneous-move game. Each player chooses their entire strategy vector $\boldsymbol{a}_n$ at the outset. Therefore, the solution concept we analyze throughout this paper is the *open-loop Nash equilibrium*.

Our non-cooperative analysis focuses on characterizing the pure-strategy Nash equilibrium of this game.

Definition 2 (Open-Loop NE)**.** *A strategy profile $\boldsymbol{a}^* = (\boldsymbol{a}_1^*, \ldots, \boldsymbol{a}_N^*)$ constitutes a pure-strategy Nash Equilibrium (NE) if, for every player $n \in \mathcal{N}$, the strategy $\boldsymbol{a}^*$ is a best response to the other players' equilibrium strategies $\boldsymbol{a}_{-n}^* = (\boldsymbol{a}_1^*, \ldots, \boldsymbol{a}_{n-1}^*, \boldsymbol{a}_{n+1}^*, \ldots, \boldsymbol{a}_N^*)$. That is:*

$$\forall n \in \mathcal{N}, \quad \forall \boldsymbol{a}_n \in \mathcal{A}_n^T, \quad U_n(\boldsymbol{a}_n^*, \boldsymbol{a}_{-n}^*; \boldsymbol{x}_1) \geq U_n(\boldsymbol{a}_n, \boldsymbol{a}_{-n}^*; \boldsymbol{x}_1).$$

For clarity, we will start by analyzing the one-stage version of this game, which corresponds to the case where the planning horizon $T = 1$.

2.4. Comments on the chosen game model and solution concept

The choice we have made here is to model the problem of interdependent emission decisions as a one-shot game model. The rationale behind this modeling choice is fourfold: (i). Discrete-time dynamic game models, such as repeated games, Markov games, or time-difference games, are built upon the one-shot game (also called a stage or component game). Analyzing the one-shot model is, therefore, a critical first step; (ii). While dynamic games are mathematically more general, they are not necessarily more accurate for this context. In terms of governmental decision-making, one-shot models often better reflect human behavior; a similar approach was utilized in modeling the Covid-19 pandemic (see, e.g., [57]), where government policy is modeled as a sequence of updated one-shot games rather than a single, long-term and complex dynamic game; (iii). Computing all feasible strategies for a dynamic game is typically a hard problem. In contrast, the proposed model provides a high-quality approximation of the feasible utility region; (iv). The proposed model is highly tractable, making it better suited for explainability and the interpretation of results.

Regarding the construction of the one-shot game model, it achieves a favorable tradeoff between realism and tractability. For instance, while the cost term could be made more granular by incorporating an individualized damage function D_n, doing so would cause the game to lose its weighted potential game structure. This specific structure is advantageous because it guarantees the existence of a pure Nash equilibrium regardless of the specific assumptions made regarding B_n, D, or the underlying system dynamics. In contrast, related works often rely on ad hoc or conditional equilibrium analyses that only hold under restrictive parameters. Furthermore, this structure facilitates computation, as iterative procedures like sequential best-response dynamics are guaranteed to converge. Adopting the Nash Equilibrium as the primary solution concept is appropriate, as sovereign nations ultimately make decisions driven by their own self-interest and autonomy. The NE constitutes, by definition, a self-enforcing agreement where no country has a unilateral incentive to deviate.

3. Non-Cooperative Equilibrium Analysis

Having defined the proposed CO_2 emission game, we now turn to the analysis of its non-cooperative outcome. The dynamic and nonlinear nature of the climate-economic relation would typically render a full characterization of the NE intractable. However, we show that our game possesses an underlying structure: it is a weighted potential game [38]. This allows a complete analysis, from guaranteeing existence to characterizing the equilibrium conditions.

3.1. The Potential Game Structure and Existence of Nash Equilibrium

We first establish the potential game property, which proves the existence of a pure-strategy NE. We begin with the one-stage ($T = 1$) case for analytical clarity, then extend the result to the full planning game ($T > 1$).

The One-stage Game ($T = 1$). The strategic interdependence in the one-stage game is captured entirely through the single global temperature variable θ^{at}. This feature implies a strong structure, as shown in the following proposition.

Proposition 1 (Potential Structure of the one-stage game)**.** *The static climate policy game Γ_1 is a weighted potential game with weights w_n. A potential function $\phi : \mathcal{A} \to \mathbb{R}$ is given by:*

$$\phi(\boldsymbol{a}; \boldsymbol{x}_1) = \sum_{n=1}^{N} \frac{B_n(a_n)}{w_n} - D(\theta^{\text{at}}(\boldsymbol{a}; \boldsymbol{x}_1)). \tag{5}$$

Proof. The proof follows directly from the definition of a weighted potential game [38]. For any player n and any two distinct actions $a_n, \tilde{a}_n \in \mathcal{A}_n$, the change in utility, $u_n(a_n, a_{-n}) - u_n(\tilde{a}_n, a_{-n})$, can be checked to be equal to $w_n[\phi(a_n, a_{-n}) - \phi(\tilde{a}_n, a_{-n})]$. □

The Planning Game ($T > 1$). Extending this result to the planning game Γ_T requires showing that the potential structure is preserved over the entire time horizon as the state of the system, $\boldsymbol{x}_t$, evolves. This is possible since the instantaneous game possesses a property known as state-transitivity.

Definition 3 (State-Transitive Weighted Potential Game [40])**.** *An instantaneous game with utility functions $u_n(\boldsymbol{a}_t; \boldsymbol{x}_t)$ is a state-transitive weighted potential game if there exists a function $\phi(\boldsymbol{a}_t; \boldsymbol{x}_t)$ and positive weights ω_n such that for any action profile $\boldsymbol{a}_t$ and any two states $\boldsymbol{x}_t, \tilde{\boldsymbol{x}}_t$, the following holds for all $n \in \mathcal{N}$:*

$$u_n(\boldsymbol{a}_t; \boldsymbol{x}_t) - u_n(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t) = \omega_n[\phi(\boldsymbol{a}_t; \boldsymbol{x}_t) - \phi(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t)].$$

Proposition 2 (State-Transitivity of the Instantaneous Game)**.** *For any period t, the instantaneous game with utility functions $u_{nt}(\boldsymbol{a}_t; \boldsymbol{x}_t)$ is a state-transitive weighted potential game with weights w_n and potential function $\phi(\boldsymbol{a}_t; \boldsymbol{x}_t)$ as defined in* (5)*.*

Proof. For any $\boldsymbol{a}_t, \boldsymbol{x}_t, \tilde{\boldsymbol{x}}_t$, the benefit term $B_n(a_{nt})$ is independent of the state. Thus, the change in utility depends only on the change in damages: $u_{nt}(\boldsymbol{a}_t; \boldsymbol{x}_t) - u_{nt}(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t) = -w_n[D(\theta^{\text{at}}(\boldsymbol{a}_t; \boldsymbol{x}_t)) - D(\theta^{\text{at}}(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t))]$. Similarly, the change in the potential function is $\phi(\boldsymbol{a}_t; \boldsymbol{x}_t) - \phi(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t) = -[D(\theta^{\text{at}}(\boldsymbol{a}_t; \boldsymbol{x}_t)) - D(\theta^{\text{at}}(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t))]$. The result follows immediately by multiplying by w_n. □

The state-transitive property of the instantaneous game allows us to construct a potential function for the full planning game, as the potential accumulates additively over time.

Proposition 3 (Potential Structure of the Planning Game)**.** *The open-loop dynamic game, i.e., the planning game Γ_T, is a weighted potential game with weights w_n. The potential function $\Phi : \mathcal{A}^T \to \mathbb{R}$ is given by the sum of the instantaneous potentials, $\phi(\boldsymbol{a}_t; \boldsymbol{x}_t)$, over the entire time horizon:*

$$\Phi(\boldsymbol{a}; \boldsymbol{x}_1) = \sum_{t=1}^{T} \phi(\boldsymbol{a}_t; \boldsymbol{x}_t), \tag{6}$$

where $\boldsymbol{x}_t$ is the state determined by the action history $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{t-1})$ from $\boldsymbol{x}_1$.

Proof. The proof combines the properties of a weighted potential game [39] and a state-transitive weighted potential game [40]. Let $\{\boldsymbol{x}_t\}_{t=1}^T$ be the state trajectory that results from the strategy profile $\boldsymbol{a}$, starting from $\boldsymbol{x}_1$. Similarly, let $\{\tilde{\boldsymbol{x}}_t\}_{t=1}^T$ be the state trajectory resulting from $\tilde{\boldsymbol{a}} = (\tilde{\boldsymbol{a}}_n, \boldsymbol{a}_{-n})$, also starting from the same initial state $\boldsymbol{x}_1$. For any period t, let $\boldsymbol{a}_t$ and $\tilde{\boldsymbol{a}}_t = (\tilde{a}_{nt}, a_{-nt})$ denote the corresponding action profiles within that period. Then, the result follows from:

$$
\begin{aligned}
&U_n(\boldsymbol{a}; \boldsymbol{x}_1) - U_n(\tilde{\boldsymbol{a}}; \boldsymbol{x}_1) = \sum_{t=1}^{T} u_{nt}(\boldsymbol{a}_t; \boldsymbol{x}_t) - \sum_{t=1}^{T} u_{nt}(\tilde{\boldsymbol{a}}_t; \tilde{\boldsymbol{x}}_t) \\
&= \sum_{t=1}^{T} u_{nt}(\boldsymbol{a}_t; \boldsymbol{x}_t) - u_{nt}(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t) + u_{nt}(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t) - u_{nt}(\tilde{\boldsymbol{a}}_t; \tilde{\boldsymbol{x}}_t) = \sum_{t=1}^{T} w_n \left(\phi(\boldsymbol{a}_t; \boldsymbol{x}_t) - \phi(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t) + \phi(\boldsymbol{a}_t; \tilde{\boldsymbol{x}}_t) - \phi(\tilde{\boldsymbol{a}}_t; \tilde{\boldsymbol{x}}_t)\right) \\
&= w_n \sum_{t=1}^{T} \phi(\boldsymbol{a}_t; \boldsymbol{x}_t) - \phi(\tilde{\boldsymbol{a}}_t; \tilde{\boldsymbol{x}}_t) = w_n \left(\Phi(\boldsymbol{a}; \boldsymbol{x}_1) - \Phi(\tilde{\boldsymbol{a}}; \boldsymbol{x}_1)\right).
\end{aligned}
$$

Since this holds for any player $n \in \mathcal{N}$ and any unilateral deviation from $\boldsymbol{a}_n$ to $\tilde{\boldsymbol{a}}_n$, the planning game Γ_T satisfies the definition of a weighted potential game. □

The existence of a potential function for the planning game is an important result. It directly implies the existence of a pure-strategy equilibrium, which is the main result of this subsection.

Proposition 4 (Existence of Nash Equilibrium)**.** *Under Assumptions 1 and 2, the planning climate policy game Γ_T has at least one pure strategy NE.*

Proof. The action space for each player, $\mathcal{A}_n^T$, is a product of compact sets and is therefore compact. The full action space $\mathcal{A}^T = \prod_n \mathcal{A}_n^T$ is also compact. The benefit and damage functions are continuous by assumption, and the geophysical state-transition functions are continuous. Therefore, the total utility function U_n is continuous for all n, and the planning potential function $\Phi(\boldsymbol{a}; \boldsymbol{x}_1)$ is a continuous function over the compact set $\mathcal{A}^T$. By the Extreme Value Theorem, Φ must attain a global maximum on $\mathcal{A}^T$. Any such maximizer is a pure-strategy Nash Equilibrium of Γ_T. □

Remark 2. *The sufficient condition to ensure the existence of at least one pure NE is the continuity of both the benefit functions and the damage function.*

4. Equilibrium Properties: Structure and Uniqueness

Having established the existence of an NE, we now investigate its fundamental properties. This section begins by characterizing the equilibrium conditions. We then utilize this framework to analyze the structural ordering of the equilibrium set and, finally, to establish sufficient conditions for uniqueness.

4.1. Characterization of the Nash Equilibrium

The economic trade-off at the center of the non-cooperative equilibrium is best understood by first examining the case of an interior solution, i.e., $a_{nt} \in (e_n^{\min}, e_n^{\max})$ for all $n \in \mathcal{N}$ and $t \in \mathcal{T}$. Here, each country freely chooses its emissions level to balance the marginal benefits and marginal costs. In particular, exhibiting conditions under which the equilibrium action is $e_n^{\max}$ is relevant since it allows us to better understand why governments do not try to decrease their CO_2 emissions.

Proposition 5 (First-Order Conditions for an Interior Equilibrium)**.** *Let $\boldsymbol{a}^*$ be an interior Nash Equilibrium of the planning game Γ_T. Then for each player n and each period $t \in \mathcal{T}$, the following necessary condition must hold:*

$$
B_n'(a_{nt}^*) = w_n \mathcal{M}_{nt}(\boldsymbol{a}^*), \tag{7}
$$

where the private marginal cost of emissions, $\mathcal{M}_{nt}(\boldsymbol{a}^*)$, is the cumulative sum of all future marginal damages caused by an additional unit of emission in period t:

$$
\mathcal{M}_{nt}(\boldsymbol{a}^*) := \sum_{s=t}^{T} \frac{\partial D(\theta_{s+1}^{\text{at}*})}{\partial a_{nt}} = \sum_{s=t}^{T} \sum_{r=s}^{T} \left(\frac{D'(\theta_{r+1}^{\text{at}*})}{\ln(2)} \cdot \frac{\boldsymbol{d}_\theta^\top \boldsymbol{A}_\theta^{r-s} \boldsymbol{b}_\theta F_{2\times CO_2}}{C_{s+1}^{\text{at}*}} \cdot \boldsymbol{d}_C^\top \boldsymbol{A}_C^{s-t} \boldsymbol{b}_C \right), \tag{8}
$$

where the introduced parameters are given in Appendix A.1.

The above condition provides the core economic intuition: in equilibrium, each country pollutes up to the point where its own marginal benefit equals its own private marginal cost, ignoring the costs inflicted upon other nations.

While intuitive, this condition only holds for interior solutions. The complete characterization of any Nash Equilibrium, including those at the boundaries of the action space, is given by the more general Karush-Kuhn-Tucker (KKT) conditions.

Proposition 6 (General Equilibrium KKT Conditions)**.** *A strategy profile $\boldsymbol{a}^*$ is a Nash Equilibrium of the planning game Γ_T only if for each player $n \in \mathcal{N}$ and each period $t \in \mathcal{T}$, there exist Lagrange multipliers $\overline{\lambda}_{nt} \geq 0$ and $\underline{\lambda}_{nt} \geq 0$ such that:*

(i) ***Stationarity:***

$$B_n'(a_{nt}^*) - w_n \mathcal{M}_{nt}(\boldsymbol{a}^*) + \underline{\lambda}_{nt} - \overline{\lambda}_{nt} = 0.$$

(ii) ***Complementary Slackness:***

$$\underline{\lambda}_{nt}(a_{nt}^* - e_n^{\min}) = 0 \quad \textit{and} \quad \overline{\lambda}_{nt}(e_n^{\max} - a_{nt}^*) = 0.$$

Proof. The conditions are the standard necessary conditions for the solution to each player n's constrained optimization problem, holding the strategies of other players $\boldsymbol{a}_{-n}^*$ fixed. The term $\mathcal{M}_{nt}(\boldsymbol{a}^*)$ is derived from the Lagrangian as shown in Appendix B.1. □

The KKT conditions include the interior solution as a special case: for an interior equilibrium ($e_n^{\min} < a_{nt}^* < e_n^{\max}$), both multipliers must be zero, and the stationarity condition collapses to the simple equality in (7). The multipliers themselves have a clear economic interpretation: $\overline{\lambda}_{nt}$ is the shadow price of the upper emissions constraint, representing the net marginal benefit the player would gain from being allowed to emit more. It is positive only when the player is at a boundary equilibrium where $a_{nt}^* = e_n^{\max}$.

4.2. Boundary Equilibria and the Role of Economic Development

The KKT framework provides a tool through which to analyze the economic drivers of different equilibrium outcomes. The possibility of a boundary solution, where $a_{nt}^* = e_n^{\max}$, is not merely a technicality but a formal representation of an economic and political reality.

As discussed in the previous section, the Lagrange multiplier $\overline{\lambda}_{nt}$ can be interpreted as the shadow value of the emissions constraint. When $\overline{\lambda}_{nt} > 0$, it signifies that the player faces a *development constraint*: their marginal benefit from emissions is so high that it remains above their private marginal cost even at their maximum possible emission level. This situation represents an industrializing nation where emissions are tightly coupled with economic growth and development. For such a player, the best response may be a dominant strategy of maximum emissions, rendering the strategic element of abatement trivial.

Conversely, a mature or post-industrial economy is more likely to have a marginal benefit curve that declines sharply, allowing for an interior solution where the multipliers are zero and $B_n'(a_{nt}^*) = w_n \mathcal{M}_{nt}(\boldsymbol{a}^*)$. This represents a larger strategic environment where countries actively weigh the costs and benefits of abatement at the margin.

To ensure our analysis focuses on this strategic interaction, we impose the following condition, which rules out the existence of dominant strategies and ensures that an interior equilibrium is at least possible for all players.

Assumption 3 (Condition for Non-Trivial Strategic Choice)**.** *For all players n and periods t, the marginal benefit at the maximum emission level is sufficiently low that it does not exceed the lowest possible private marginal cost a player could face:*

$$B_n'(e_n^{\max}) < \min_{\boldsymbol{a}_{-nt} \in \mathcal{A}_{-nt}^T} \left\{ w_n \mathcal{M}_{nt}(e_n^{\max}, \boldsymbol{a}_{-nt}) \right\}.$$

This assumption allows us to focus on the central strategic trade-offs of the planning game, ensuring that far-sighted players face a non-trivial choice between development and abatement. However, we note that this condition depends on the full, cumulative marginal cost $\mathcal{M}_{nt}$. In the case of myopic agents (which we analyze as a benchmark in Section 6), players ignore future damages, effectively perceiving a much lower marginal cost. Consequently, myopic players may fail to satisfy this condition and revert to a dominant strategy of maximum emissions ($a_{nt} = e_n^{max}$), a behavior that correctly characterizes the "business-as-usual" tragedy of the commons.

4.3. Structural Ordering of Nash Equilibria

Beyond characterizing a single equilibrium, we can make a statement about the structure of the entire equilibrium set. The strict concavity of the benefit functions imposes a strong ordering on any potential multiple NE. The following proposition shows that if two NE differ, they must be ordered in a specific way across all players within a given time period.

Proposition 7 (Nash Equilibrium Ordering)**.** *Assume that for all $n \in \mathcal{N}$ and $t \in \mathcal{T}$, B_n is strictly concave. Let $\boldsymbol{a}^*$ and $\tilde{\boldsymbol{a}}^*$ be two distinct NE of Γ_T. If for some period t and some player n, their emissions differ (e.g., $a_{nt} < \tilde{a}_{nt}$), then for that same period t, all other players' emissions must be weakly ordered in the same direction, i.e., $a_{kt} \leq \tilde{a}_{kt}$ for all $k \in \mathcal{N}$. Furthermore, equality ($a_{nt} = \tilde{a}_{nt}$) can only hold if the action of player n is at a boundary of its action set, $a_{nt} \in \{e_n^{\min}, e_n^{\max}\}$.*

Proof. The proof proceeds by contradiction. Assume for some period t, there exist players n and k such that $a_{nt} < \tilde{a}_{nt}$ and $a_{kt} > \tilde{a}_{kt}$. We use the planning first-order conditions from Proposition 6 for both players under both equilibria. The strict concavity of B_n and B_k leads to a contradiction, as it implies both $a_{kt} \leq \tilde{a}_{kt}$ and $a_{kt} > \tilde{a}_{kt}$. The argument is presented in Appendix B.2. □

This result rules out equilibria that are *tangled*, i.e., equilibria where some players' emissions are not ordered across the two equilibria. It implies that if multiple equilibria exist, they can be Pareto-ordered in terms of per-period emissions. This structure is necessary for understanding the potential for coordination failure and is a base to analyzing uniqueness.

4.4. Sufficient Conditions for Uniqueness

While Proposition 7 structures the NE set, it does not guarantee NE uniqueness, which facilitates the game analysis, e.g., for predicting the game outcome. We now provide sufficient conditions that collapse the set to a unique point. These conditions place restrictions on the relative curvatures of the benefit and damage functions. For clarity, we present first some conditions for uniqueness in the one-stage game and then provide an extended one for the planning game.

Proposition 8 (Sufficient Conditions for Uniqueness in the one-stage game)**.** *Under Assumptions 1 and 2, the one-stage game Γ_1 has a unique NE if any of the following sufficient conditions hold for all feasible action profiles:*

(i) ***Rapidly Accelerating Damages:*** *The elasticity of the marginal damage function with respect to atmospheric carbon concentration is greater than one. Formally, for any two concentration levels $C^{\text{at}} < \tilde{C}^{\text{at}}$:*

$$\frac{D'(\theta^{\text{at}}(\widetilde{C}^{\text{at}}))}{D'(\theta^{\text{at}}(C^{\text{at}}))} \geq \frac{\widetilde{C}^{\text{at}}}{C^{\text{at}}}.$$

(ii) ***Rapidly Diminishing Benefits:*** *There exists at least one player n whose marginal benefits diminish faster than the inverse of the carbon concentration. Formally, for any two action profiles $\boldsymbol{a}, \tilde{\boldsymbol{a}}$ where $a_n < \tilde{a}_n$:*

$$\frac{B_n'(\tilde{a}_n)}{B_n'(a_n)} \leq \frac{C^{\text{at}}(\boldsymbol{a})}{C^{\text{at}}(\tilde{\boldsymbol{a}})}.$$

(iii) ***Sufficiently Exponential Damage Function:*** *The potential function $\phi(\boldsymbol{a})$ is strictly concave in $\boldsymbol{a}$. A sufficient condition for this is:*

$$D''(\theta^{\text{at}}) \geq \frac{D'(\theta^{\text{at}})}{\beta_\theta}.$$

Proof. (i) and (ii) are proved by contradiction. Assuming two distinct equilibria and using the ordering property from Proposition 7 and the first-order conditions from Proposition 5 leads to a violation of the stated condition. (iii) follows from the property that a strictly concave potential function has a unique maximizer, which corresponds to a unique Nash Equilibrium. The proofs are in the Appendix B.3. □

While the static case identifies several independent paths to uniqueness, not all of them generalize to the planning game. Specifically, the inter-temporal summation in the marginal cost term complicates the comparison of first-order conditions. However, the structural properties of the game (those based on its strategic interaction or the curvature of its potential function) provide foundations for proving uniqueness in the full planning game Γ_T.

Proposition 9 (Sufficient Conditions for Uniqueness in the Planning Game)**.** *Under Assumptions 1 and 2, the planning game Γ_T has a unique Pure-Strategy NE if either of the following conditions holds for all feasible action profiles and resulting state trajectories:*

(i) ***Strategic Substitutes:*** *For any period $s \in \mathcal{T}$, the marginal damage per unit of carbon,*

$$g_s(\boldsymbol{E}) = \frac{D'(\theta^{\text{at}}_{s+1}(\boldsymbol{E}))}{C^{\text{at}}_{s+1}(\boldsymbol{E})},$$

is a non-decreasing function of aggregate emissions in any period $u \leq s$.

(ii) ***Sufficiently Exponential Damage Function:*** *The planning potential function $\Phi(\boldsymbol{a})$ is strictly concave in the full action profile $\boldsymbol{a}$. A sufficient condition for this is that for any reachable state:*

$$D''(\theta^{\text{at}}) \geq \frac{D'(\theta^{\text{at}})}{\beta_\theta}.$$

Proof. For (i), the proof first establishes that the stated condition on the marginal damage function $g_s(\boldsymbol{E})$ ensures the game has strategic substitutes. We then use this structural property in a proof by contradiction. Assuming the existence of two distinct NE and comparing their first-order conditions leads to a violation of the strict concavity of the benefit functions. Therefore, the equilibrium must be unique.

For (ii), the proof follows from the same argument in the static case Γ_1 by summing the potential function at each state, which is concave under this assumption.

The detailed proofs are in Appendix B.4. □

The significance of Proposition 9 lies in its unification of the mathematical condition with the game's core strategic nature. In contrast to the static analysis (in the *One-stage* case), which offers a collection of "unrelated" technical assumptions, the dynamic result demonstrates that uniqueness is a direct consequence of the problem's structure as a game of *strategic substitutes*. This property implies that as opponents increase their emissions over time, the marginal cost of any new emission rises for all players. It is this strategic feedback, where aggressive play by others diminishes one's own incentive to pollute, that tames the multi-period interaction to a unique non-cooperative trajectory.

5. Inefficiency and Strategic Stability of International Environmental Agreements

Having characterized the fully non-cooperative equilibrium, with the solution concept of NE, we now turn to the central challenge of international climate policy: achieving cooperation. This section proceeds by first defining the critical benchmark to the non-cooperative solution: the fully cooperative Social Optimum (SO). We then demonstrate the tension at the root of the problem by proving the inherent strategic instability of full cooperation. Finally, we characterize the inefficiency of the NE solution concept, thereby formalizing the core paradox that motivates the study of partial agreements.

5.1. Analytical Benchmarks: Non-Cooperation and the Social Optimum

To analyze the problem of cooperation, we compare the outcomes of a game played under two opposite paradigms, following the canonical IEA literature [5].

- The **non-cooperative outcome** corresponds to a situation where each country acts purely in its self-interest. The resulting strategy is the open-loop Nash Equilibrium $\boldsymbol{a}^*$, as derived in the preceding sections.

- The **socially optimal outcome**, or first-best, corresponds to a hypothetical situation where all countries form a *Grand Coalition* and act to maximize their collective welfare. The resulting emissions profile, denoted by $\boldsymbol{a}^{\mathrm{SO}}$, is the solution to:

$$\max_{\boldsymbol{a}\in\mathcal{A}^T} W(\boldsymbol{a};\boldsymbol{x}_1) := \sum_{n=1}^{N} U_n(\boldsymbol{a};\boldsymbol{x}_1), \tag{9}$$

  where W is the total welfare of all the agents.

The first-order conditions for the SO reveal its rigid nature: each player's marginal benefit from emissions must be equated not to their private marginal cost, but to the full *global* marginal social cost.

Proposition 10 (First-Order Conditions for the Social Optimum)**.** *Let $\boldsymbol{a}^{\mathrm{SO}}$ be an interior social optimum. Then for each player n and each period t, the following condition must hold:*

$$B_n'(a_{nt}^{\mathrm{SO}}) = \left(\sum_{k=1}^{N} w_k\right) \sum_{s=t}^{T} \frac{\partial D(\theta_{s+1}^{\mathrm{at,SO}})}{\partial a_{nt}} = \left(\sum_{k=1}^{N} w_k\right) \mathcal{M}_{nt}(\boldsymbol{a}^{\mathrm{SO}};\boldsymbol{x}_1). \tag{10}$$

Proof. The proof follows from deriving the KKT conditions for the constrained optimization problem in (9). The derivation is analogous to that in Proposition 6 and is omitted for brevity. □

Similar conditions to Proposition 6 and 10 are obtained for the non-interior SO. However, we will see in the next section that non-interior SO are particular cases where SO and NE overlap.

5.2. Strategic Instability of Full Cooperation

The core strategic tension of climate policy is revealed by evaluating a single country's incentive at the socially optimal profile. While $\boldsymbol{a}^{\mathrm{SO}}$ is best for the collective, it is not individually rational for any single player, giving rise to a strong incentive to free-ride.

Proposition 11 (Incentive to Defect from the Social Optimum)**.** *At an interior Social Optimum $\boldsymbol{a}^{\mathrm{SO}}$, every player $n \in \mathcal{N}$ has a strict incentive to unilaterally increase their emissions. That is, the gradient of player n's utility function, evaluated at $\boldsymbol{a}^{\mathrm{SO}}$, is strictly positive in all components:*

$$\left.\frac{\partial U_n(\boldsymbol{a};\boldsymbol{x}_1)}{\partial a_{nt}}\right|_{\boldsymbol{a}=\boldsymbol{a}^{\mathrm{SO}}} > 0 \quad \forall t \in \mathcal{T}.$$

Proof. A detailed proof is provided in Appendix C.1. □

This incentive to defect renders full cooperation strategically not rational. To formalize this conclusion, we employ the standard model of IEA formation, which proceeds as a two-stage game [5].

i) **Stage 1: The Membership Game.** Each country $n \in \mathcal{N}$ simultaneously decides whether to join the climate coalition (become a signatory) or not (remain a non-signatory). This results in a coalition $C \subseteq \mathcal{N}$.

ii) **Stage 2: The Emissions Game.** Given the coalition structure C, all countries choose their emissions profiles. Signatories act cooperatively to maximize their *joint* welfare, while non-signatories act non-cooperatively to maximize their *individual* welfare, with each group taking the actions of the other as given.

The solution concept is a subgame perfect equilibrium. An agreement is considered self-enforcing, or **stable**, if no single country has an incentive to change its membership status.

Definition 4 (Coalition Stability)**.** *Let C be a coalition of signatories. The coalition C is stable if and only if it satisfies:*

*i) **Internal Stability:** No signatory $n \in C$ can increase its utility by unilaterally leaving the coalition.*

$$U_n(\boldsymbol{a}^*(C)) \geq U_n(\boldsymbol{a}^*(C \setminus \{n\})) \quad \forall n \in C,$$

where $\boldsymbol{a}^*(C)$ is the NE for the coalition C.

ii) ***External Stability:*** *No non-signatory $j \notin C$ can increase its utility by unilaterally joining the coalition.*

$$U_j(\boldsymbol{a}^*(C)) \geq U_j(\boldsymbol{a}^*(C \cup \{j\})) \quad \forall j \notin C.$$

With this framework, the implication of the free-rider incentive becomes a formal result: the Grand Coalition is inherently unstable.

Corollary 1 (The Inherent Instability of the Grand Coalition)**.** *The Grand Coalition $C = \mathcal{N}$ is not internally stable.*

Proof. This is a direct consequence of Proposition 11, which proves that every member of the Grand Coalition has a strict incentive to defect from the cooperative profile $\boldsymbol{a}^{\mathrm{SO}} = \boldsymbol{a}^*(\mathcal{N})$. The utility from defecting is thus strictly greater than the utility from cooperating, violating the condition for internal stability. □

5.3. Inefficiency of Non-Cooperation

Having established that the SO solution concept is strategically unstable[3], we now compare it with the NE strategy. The misalignment of incentives ensures that the NE is not only different from the social optimum but also results in unambiguous over-pollution under reasonable conditions.

Proposition 12 (Non efficiency of interior NE)**.** *An interior Nash Equilibrium $\boldsymbol{a}^*$ cannot be an SO.*

Proof. Assume for contradiction that an interior Nash Equilibrium $\boldsymbol{a}^*$ is also a Social Optimum. Then for any (n, t), the first-order conditions for both must hold simultaneously:

$$B'_n(a^*_{nt}) = w_n \mathcal{M}_{nt}(\boldsymbol{a}^*) \quad \text{and} \quad B'_n(a^*_{nt}) = \left(\sum_{k=1}^{N} w_k\right) \mathcal{M}_{nt}(\boldsymbol{a}^*)$$

Equating the right-hand sides yields $w_n \mathcal{M}_{nt}(\boldsymbol{a}^*) = (\sum_{k=1}^{N} w_k)\mathcal{M}_{nt}(\boldsymbol{a}^*)$, which simplifies to $(\sum_{k\neq n} w_k)\mathcal{M}_{nt}(\boldsymbol{a}^*) = 0$. Since $w_k > 0$ and the marginal damage term $\mathcal{M}_{nt}(\boldsymbol{a}^*)$ is strictly positive for an interior solution, this is a contradiction. Thus, an interior NE cannot be a Social Optimum. □

Proposition 13 (Over-Pollution in the Nash Equilibrium)**.** *If the damage function satisfies conditions that guarantee a unique equilibrium (e.g., Proposition 9-(ii)), then emissions in the Nash Equilibrium are inefficiently high compared to the social optimum, i.e., $\boldsymbol{a}^* > \boldsymbol{a}^{\mathrm{SO}}$ component-wise.*

Proof. A formal proof is provided in Appendix C.2. The logic relies on the strict concavity of each player's utility function. The strictly positive gradient at the social optimum $\boldsymbol{a}^{\mathrm{SO}}$ (from Proposition 11) ensures that the optimal response to cooperation involves higher emissions, pushing the non-cooperative equilibrium towards over-pollution. □

This analysis confirms the central paradox of cooperation within our dynamic framework: the attracting outcome ($\boldsymbol{a}^*$) is inefficient, and the efficient outcome ($\boldsymbol{a}^{\mathrm{SO}}$) is unstable. We observe and quantify this inefficiency in the following numerical section.

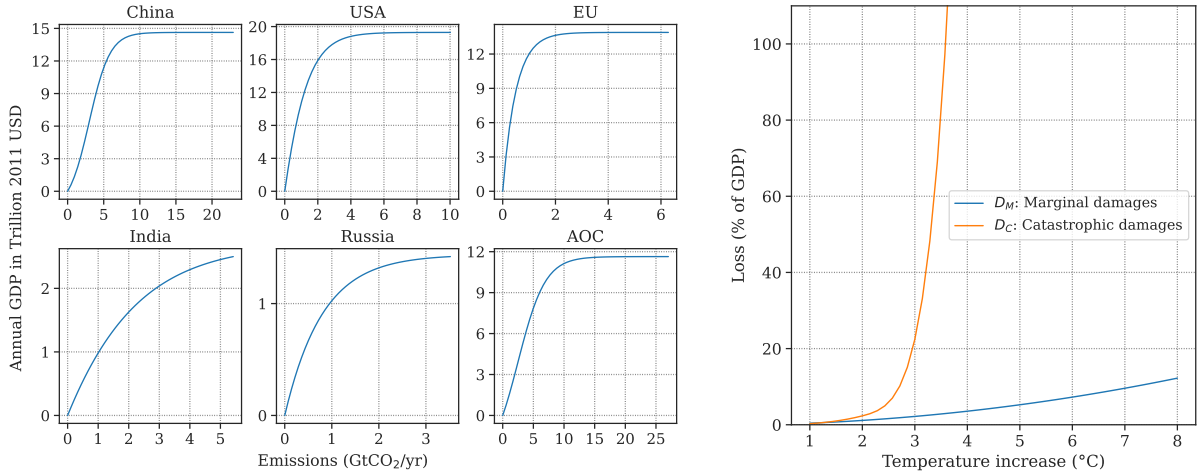
6. Numerical Analysis

Having established the theoretical bases of our game, we now proceed with a numerical analysis to explore its quantitative properties. We simulate the game for a set of $N = 6$ key players: China, USA, EU, India, Russia, and a "All Other Countries" (AOC). Those players will aggregate their utility over a time horizon to the year 2100 (corresponding to $T = 16$ five-year periods). The objective is to quantify the outcomes under different solution concepts and, most importantly, to explore the conditions under which meaningful cooperation can emerge. All simulations were implemented in Python, with the replication code available at [61].

[3] Note that the considered cooperative point might become an NE by introducing mechanisms such as pricing or by reformulating the considered problem as a dynamic game.

Player	$e_n^{\min}$	$e_n^{\max}$	$\text{GDP}_n^{\max}$	$\widetilde{w}_n$	r_n	e_n	$\text{GDP}_n^{\%\text{green}}$
China	0	23	146.3	1.1847	1.15	2	0
USA	0	10	192.9	1.1941	1.6	-2	0
EU	0	6.3	138.9	1.1248	2	-5	0
India	0	5.4	25	0.9074	0.5	-4	0
Russia	0	3.5	14.2	1.2866	1.25	-4	0
AOC	0	27	116.4	1.1847	0.75	2.5	0

Table 1: Player parameters used in the numerical analysis, calibrated from [58, 59]. Emissions bounds ($e_n^{\min}$, $e_n^{\max}$) are in $GtCO_2$/yr and are taken as twice the current emission (from [58]) to take into account potential emission increase. Max GDP ($\text{GDP}_n^{\max}$) is in 10^{11} US$ (2011). Vulnerability weights (w_n) are from [60].



(a) Heterogeneous Player Incentives: Calibrated Benefit Functions.

(b) Two Strategic Worlds: Calibrated Damage Functions.

Figure 1: **Core components of the game-theoretic model. (a)** The calibrated benefit functions for the six players/regions, based on (11). The diverse shapes, from the sharply concave, post-industrial profiles of the USA and EU to the extended growth profile of India, represent the heterogeneous economic incentives and development stages that drive strategic behavior. **(b)** The two archetypal damage functions that define the strategic worlds of the analysis. The Marginal Damages function (12) models a conventional world of smoothly increasing costs, while the *Catastrophic damage* function (13) models a world with a *Tipping point* dynamic, creating fundamentally different strategic scenarios.

6.1. Model Calibration and Experimental Design

To illustrate our theoretical framework, we must specify functional forms and parameters for the players' utility functions and the underlying geophysical model. These choices represent crucial assumptions about economic realities and climate damage that define the strategic environment of the game. Our experimental design is structured along two primary dimensions: first, the perceived severity of the climate threat (the "state of the world"); and second, the institutional protocol for decision-making (the "rules of the game").

Benefit Functions: Modeling Heterogeneous Economies. To operationalize the general benefit function $B_n(a_{nt})$ from Assumption 1, we require a specification that is flexible enough to capture the diverse economic structures of the players. We adopt a sigmoidal (logistic) functional form, which can represent the full lifecycle of an economy's emissions-growth relationship. The specific form is:

$$B_n(a_n) = \frac{\text{GDP}_n^{\max}\left(1 - \text{GDP}_n^{\%\text{green}}\right)}{f_n(e_n^{\max}) - f_n(e_n^{\min})}\left(f_n(a_n) - f_n(e_n^{\min})\right) + \text{GDP}_n^{\%\text{green}} \times \text{GDP}_n^{\max}, \tag{11}$$

with the logistic function $f_n(a_n) = (1+e^{-r_n(a_n-e_n)})^{-1}$. For simplicity and because of its prevalence in government policy, we consider a GDP-based benefit function. However, future iterations could incorporate more holistic social wellness metrics, such as the Genuine Progress Indicator [62], which accounts for environmental externalities, or the Human

Development Index [63], which integrates health and education outcomes. Other alternatives include the Inclusive Wealth Index [64] to better reflect natural capital depletion or subjective measures like Gross National Happiness [65].

This S-shaped curve represents the various stages of economic development. The initial increasing part represents an early-stage, industrializing economy (e.g., India in our model) where emissions are tightly and increasingly coupled with rapid GDP growth [66, 67]. The final plateau part represents a mature, service-based economy (e.g., the EU or USA) that has begun to decouple emissions from growth and exhibits diminishing marginal returns to pollution.

By parameterizing this function differently for each player, as detailed in Table 1 and visualized in Figure 1a, we explicitly model the **heterogeneity** of incentives. This moves beyond the symmetric-player assumption common in early IEA models and is crucial for understanding the real-world challenges of coalition formation.

Geophysical Model Parameterization. To ensure our strategic analysis is grounded in a credible representation of the climate system, we employ an SCM parameterized to align with the projections of more complex Earth System Models used by the IPCC [24]. Specifically, we use the carbon cycle module from [36] and the temperature module from [35]. As shown in Figure 2, the model's response to standard emission scenarios (e.g., Representative Concentration Pathways (RCP) [68, 69, 70]) is consistent with established scientific benchmarks.

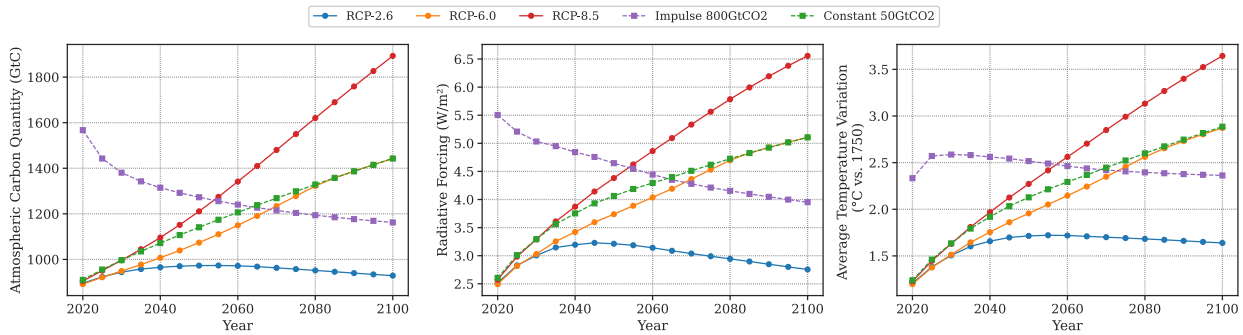


Figure 2: **Validation of the SCM against standard climate scenarios.** The three panels show the calibrated SCM's response under standard emissions trajectories, including the IPCC's Representative Concentration Pathway (RCP). The model correctly reproduces the key dynamic relationships between atmospheric carbon concentration (left), radiative forcing (middle), and global mean temperature anomaly (right). The consistency of these projections with established climate science benchmarks validates the SCM as a credible physical foundation for the strategic analysis.

Defining Two Strategic Worlds with the Damage Function. The objective of our experiment is to investigate how the strategic environment evolves in response to the perceived severity of climate-related damages. We do this by defining two distinct and archetypal worlds, each anchored by a specific damage function that represents a different understanding of climate damage under realistic SMC modeling. In both worlds, damages from the function D are calculated as a percentage of GDP, and the function is scaled by a player-specific vulnerability weight $w_n = \widetilde{w}_n \text{GDP}_n^{\max}$, where w_n is a geographical vulnerability factor from the FUND model [60].

- **The Marginal Damages World:** This scenario represents the conventional view of climate change as a standard externality with smoothly increasing costs (e.g., [71, 48, 72]). It is defined by a standard quadratic damage function, representative of forms used in many integrated assessment models:

$$D_M(\theta^{\text{at}}) = 0.16(\theta^{\text{at}})^2 + 0.25\theta^{\text{at}}. \tag{12}$$

  Strategically, this world gives rise to a game of *managing a marginal externality*, where the gains from cooperation are expected to be modest, and the classic free-rider problem is anticipated to be dominant.

- **The Catastrophic Risk World:** This scenario simulates a world with nonlinear risks and a potential "Tipping point", reflecting concerns from climate science [30, 73]. The damage function is specified as a piecewise

function that accelerates increasingly after a threshold:

$$D_C(\theta^{\text{at}}) = \begin{cases} 0.0044(\theta^{\text{at}})^2 & \text{if } \theta^{\text{at}} \leq 2°\text{C} \\ 0.0044(\theta^{\text{at}})^2 + 0.04(\theta^{\text{at}} - 2)^3 & \text{if } \theta^{\text{at}} > 2°\text{C} \end{cases} \tag{13}$$

This Catastrophic function is quadratic for warming below 2°C, then becomes dominated by a steeply convex cubic term thereafter, representing the accelerating damages that might occur after crossing a critical climate threshold. Strategically, this world gives rise to a game of *collective risk avoidance*, creating a "climate cliff" where the value of cooperation may be orders of magnitude larger.

These two specifications, which are visualized in Figure 1b, anchor our analysis and allow us to test whether the perception of climate damage is a sufficient condition to reshape strategic incentives.

Three Policy Frameworks: Myopia, Commitment, and Adaptation. Within each strategic world, we analyze the game under three distinct interaction protocols. These frameworks are designed to capture a spectrum of strategic foresight, from extreme short-termism to perfect, long-term commitment, and a more realistic adaptive alternative.

- **The Myopic Benchmark (Repeated One-stage):** In this framework, players are pathologically short-sighted. Every five years, they solve a static, one-shot game ($T = 1$), choosing their emissions to maximize only their utility in that single period. They are completely blind to the future, path-dependent consequences of their actions on the climate stock. This serves as a benchmark for myopic, non-strategic behavior.
- **The Full Commitment Framework (Planning):** This corresponds directly to the open-loop Nash Equilibrium analyzed in our theoretical sections. Players are perfectly far-sighted and solve the full dynamic game at the outset, choosing a complete emissions trajectory from 2020 to 2100 ($T = 16$) to which they credibly commit. This framework isolates the strategic incentives under the assumption of perfect, unwavering foresight and serves as a crucial theoretical benchmark.
- **The Adaptive Framework (Receding Horizon):** This framework provides a more realistic model of policy-making that blends foresight with adaptation. Players plan over a long horizon (e.g., $T = 16$), but they only implement the first five-year step of their optimal plan. At the next five-year interval, the state of the climate system has evolved, and they resolve the *entire* planning problem from that new state forward. This iterative process of planning, acting, and re-planning captures the spirit of adaptive governance. It is the game-theoretically analogous to the "ratchet mechanism" of the Paris Agreement, where commitments are periodically revisited and strengthened.

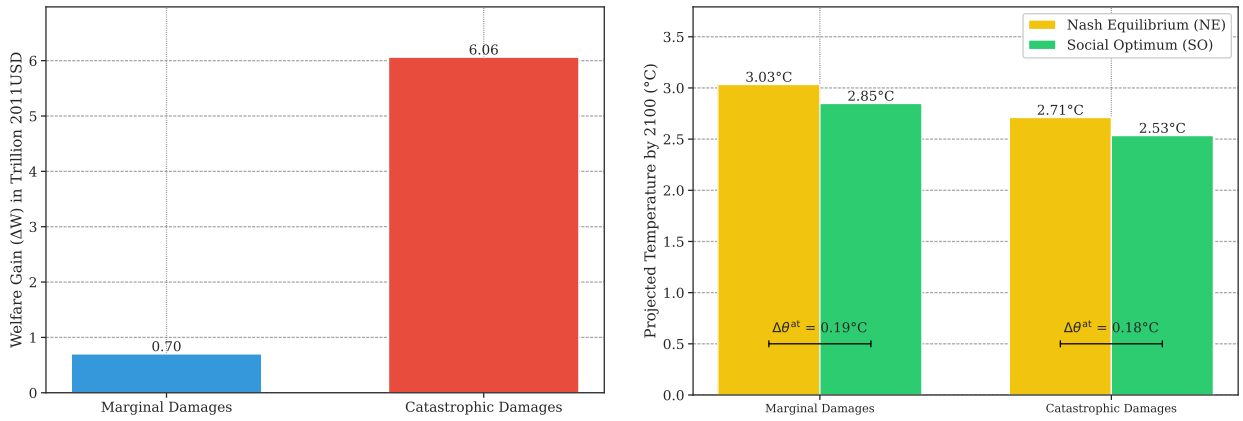
By simulating the non-cooperative and socially optimal outcomes at the intersection of these two worlds and three frameworks, we can provide a robust, quantitative assessment of the conditions under which meaningful climate cooperation can emerge.

6.2. The Endogenous Value of Cooperation

Having defined our two strategic worlds, we now test our central hypothesis: that the strategic environment of climate policy is endogenous to the severity of the damage function. To provide the cleanest test of this hypothesis, we compare NE and SO outcomes under the **Planning (full-commitment to 2100) framework**. This framework, in which players optimize their strategies over the entire horizon up to the year 2100, allows us to cleanly isolate the impact of the damage function's shape by holding the strategic time horizon fixed and common across both scenarios. Figure 3 presents the main result of this comparison.

The results illustrate, as before, a transformation of the strategies when climate change is perceived as a catastrophic threat. Panel 3a quantifies the economic value of achieving full cooperation (i.e., the formation of the Grand Coalition). To measure the gain induced by cooperation, let us define the global welfare gain as:

$$\Delta W = W^{\text{SO}} - W^{\text{NE}} := W(\boldsymbol{a}^{\text{SO}}; \boldsymbol{x}_1) - W(\boldsymbol{a}^{\text{NE}}; \boldsymbol{x}_1) = \sum_{n=1}^{N} U_n(\boldsymbol{a}^{\text{SO}}; \boldsymbol{x}_1) - \sum_{n=1}^{N} U_n(\boldsymbol{a}^{\text{NE}}; \boldsymbol{x}_1),$$



(a) Economic Value of Cooperation (ΔW)

(b) Physical Climate Outcomes

Figure 3: **The Cooperation Dividend is Endogenous to Perceived Climate Damage.** All simulations use the *Planning* framework to isolate the effect of the damage function. (a) The global welfare gain (ΔW) realized by moving from the NE to the SO. The benefit of transitioning from non-cooperative (NE) to social-optimum (SO) behavior is most pronounced in a high-risk scenario, where the welfare dividend is nearly an order of magnitude greater. (b) The absolute temperature outcomes by 2100, showing that perceived climate damage induces greater unilateral abatement even in the non-cooperative equilibrium.

where $\boldsymbol{x}_1$ is given by the model (see Appendix A) and the action profiles $\boldsymbol{a}^{\text{NE}}$ and $\boldsymbol{a}^{\text{SO}}$ are obtained numerically.

In the *Marginal Damages world*, the welfare dividend is modest, amounting to $\Delta W = \$0.70$ Trillion. This outcome reflects the classic "paradox of cooperation", where the gains from cooperation are relatively small. Thus, even with a proper climate model, cooperation provides only a weak incentive to overcome the free-rider problem. In stark contrast, the *Catastrophic Damages world* fundamentally alters this result: the welfare dividend increases by nearly an order of magnitude to $\Delta W = \$6.06$ Trillion. In this high-stakes environment, forming the Grand Coalition represents a first-order economic priority.

Panel 3b reveals the underlying physical and strategic dynamics driving this economic result. We observe two key mechanisms:

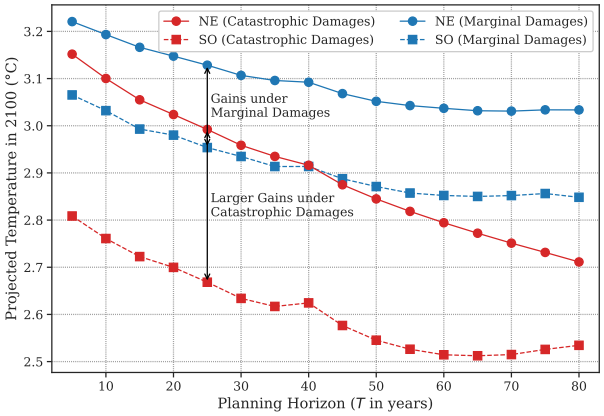
- We identify a strong effect of perceived damage on non-cooperative behavior. The Nash Equilibrium outcome is itself significantly better in the *Catastrophic Damages world* (2.71°C) than in the *Marginal Damages world* (3.03°C). This demonstrates that the mere presence of a potential tipping point disciplines even self-interested players into adopting a more cautious, lower-emissions strategy, leading to greater unilateral abatement.

- The substantial welfare dividend is driven by the high economic valuation of the cooling that cooperation provides. In both worlds, moving from the NE to the SO yields a similar amount of cooling: approximately 0.19°C in the *Marginal Damages world* and 0.18°C in the *Catastrophic Damages world*. However, in the *Catastrophic Damages world*, this modest reduction in temperature averts a region of steeply accelerating damages. It is the immense economic value of this cooling, which moves society from a dangerous outcome (2.71°C) to a significantly safer one (2.53°C), that explains the $6.06 Trillion welfare gain.

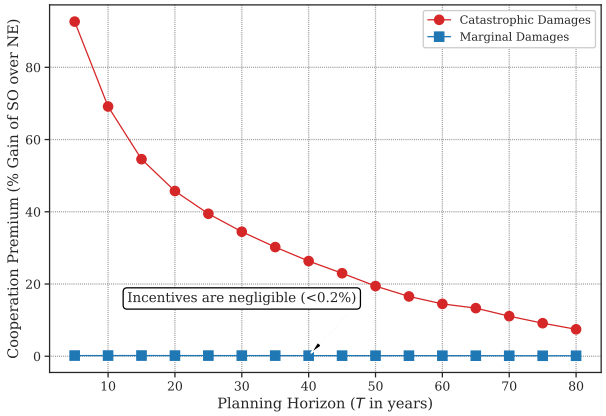
Together, these findings provide our first result: the incentive to cooperate on climate change is not a fixed parameter but is a direct consequence of the shared perception of physical risk. When the threat is understood to be nonlinear and potentially catastrophic, the value of designing effective, cooperative institutions becomes critically important. This principle, established here in a clean experimental setting, motivates our subsequent analysis of strategic foresight and adaptive policy frameworks.

6.3. Strategic Foresight and the Amplifying Effect of Damage Function

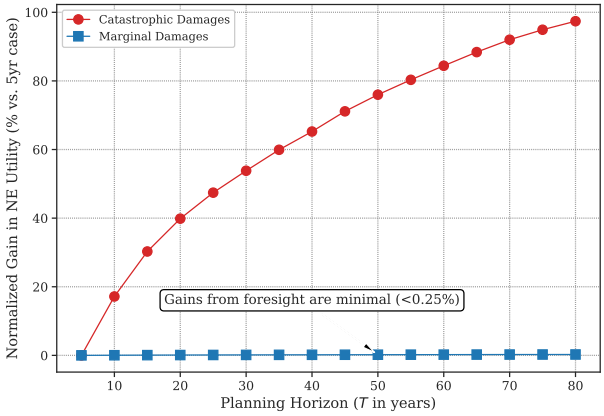
While the perception of climate damage disciplines behavior, the extent to which players act on that risk depends on their strategic foresight: the length of their planning horizon (T). We now explore this interaction.



(a) The Temperature Gains from Strategic Foresight



(b) The Economic Gains from Cooperation



(c) The Economic Gains from Strategic Foresight

Figure 4: **The Impact of Strategic Foresight (T) and Perceived Damage.** All simulations use the *Planning* framework with a 2100 horizon. (a) Projected 2100 temperature for non-cooperative (NE, solid lines) and cooperative (SO, dashed lines) outcomes. (b) The "Cooperation Premium," measured as the percentage gain in global welfare from achieving the SO over the NE. (c) The economic value of foresight, measured as the percentage gain in NE utility for a given horizon compared to a myopic one-stage ($T = 1$, i.e., 5-year) case. The plots demonstrate that perceived damage creates significant, non-negligible incentives for both foresight and cooperation. Specifically, governments have a strong incentive to adopt a long-term horizon (large T) rather than relying on a sequence of successive short-term plans (small T).

Figure 4 presents a synthesized view of the game's outcomes, demonstrating how foresight and damage perception jointly determine environmental outcomes and economic incentives when players plan to a fixed 2100 horizon.

Panel 4a illustrates the direct environmental impact of strategic foresight. As the planning horizon (T) extends, projected warming decreases across all scenarios. This mechanism is a direct consequence of our theoretical model: a longer horizon means the dynamic private marginal cost ($\mathcal{M}_{nt}$) accumulates future damages over more periods, forcing even non-cooperative players to internalize more of the externality. This effect is strongly amplified by perceived economic damage. Extending the horizon from 5 to 75 years yields only a modest ≈ 0.2°C reduction in the *Marginal Damages world*, but it delivers a substantial ≈ 0.5°C reduction in the *Catastrophic Damages world*. When the future is perceived as dangerous, far-sighted planning is a central tool for climate damage management.

Panels 4b and 4c quantify the nuanced economic incentives underpinning these physical outcomes. To ensure comparability across different planning horizons, we define $W^{\mathrm{NE}}(T)$ and $W^{\mathrm{SO}}(T)$ as the total cumulative global welfare over the full simulation period (2020–2100, i.e., 16 steps) resulting from players adopting a planning horizon of T. For cases where $T < 16$, the game is repeated sequentially until the year 2100 is reached, and the utilities are summed over the entire trajectory.

We measure the economic gain using two distinct metrics. First, the *Cooperation Premium* (Panel 4b) represents

the relative welfare gain achieved by forming the Grand Coalition compared to the non-cooperative outcome, for a fixed level of foresight T:

$$\text{Cooperation Premium } (T) = \frac{W^{\text{SO}}(T) - W^{\text{NE}}(T)}{\left|W^{\text{NE}}(T)\right|} \times 100.$$

Second, the *Value of Foresight* (Panel 4c), or the normalized gain in NE, measures the private economic incentive to extend one's planning horizon. It is calculated as the percentage increase in non-cooperative welfare gained by adopting a horizon T relative to the myopic baseline of $T = 1$ (5 years):

$$\text{Value of Foresight } (T) = \frac{W^{\text{NE}}(T) - W^{\text{NE}}(1)}{\left|W^{\text{NE}}(1)\right|} \times 100.$$

Note that we normalize by the absolute value of the baseline utility to strictly measure the magnitude of improvement, as utility values may be negative depending on the damage calibration.

In the *Marginal Damages world* (blue squares), the economic gains from both cooperation and foresight are consistently negligible, confirming that in a low-damages world, players have little incentive to either cooperate or plan for the long term.

The story is fundamentally different in the *Catastrophic Damages world* (red circles). Panel 4b shows that the Cooperation Premium, representing the welfare gain from forming the Grand Coalition, is now significant, peaking at over 6% for myopic players. This demonstrates that perceived damage creates a meaningful, non-negligible incentive to cooperate. However, this incentive diminishes as players become more far-sighted.

Conversely, Panel 4c shows that the economic value of foresight itself is also substantial under catastrophic damage. A non-cooperative player with a 75-year horizon is nearly 8% better off in utility terms than a myopic one (5-year horizon). This shows that far-sighted, self-interested behavior is not only environmentally beneficial but also economically rational.

Finally, reading these panels together reveals a sophisticated interaction. In the Catastrophic damage scenario, the premium for explicit cooperation (Panel 4b) is highest when foresight is low, and it decays as the utility gains from unilateral long-term planning (Panel 4c) increase. This leads to our second major finding: under conditions of severe climate damage, **sufficient strategic foresight can act as a partial substitute for explicit cooperation**. The shared threat becomes such an important coordinating device for far-sighted actors that it forces a convergence of selfish and socially optimal behavior, doing some of the work that a formal treaty would otherwise have to do.

6.4. The Critical Role of the Terminal Horizon in a Catastrophic Damage World

Our analysis thus far has established that perceived damage amplifies the value of both cooperation and foresight within a standard 2100 policy horizon. However, the physical impacts of 21st-century emissions will persist for centuries. The choice of a 2100 end-date, while conventional, is therefore an arbitrary and potentially misleading assumption. We now test how the economic incentives for cooperation and foresight change as we extend the terminal horizon of the game itself (T) if agents are short-sighted: if they only want to maximize over $T_{\text{end}} \leq T$ steps.

To rigorously compare these scenarios, we generalize the metrics defined in Section 6.3 to account for the variable duration of the game. Let $T_{\text{end}} \in \{2050, 2100, 2150\}$ denote the terminal year of the simulation. We define $W^{\text{NE}}(T, T_{\text{end}})$ and $W^{\text{SO}}(T, T_{\text{end}})$ as the cumulative global welfare accumulated from 2020 up to T_{end}, given a recurring planning horizon T.

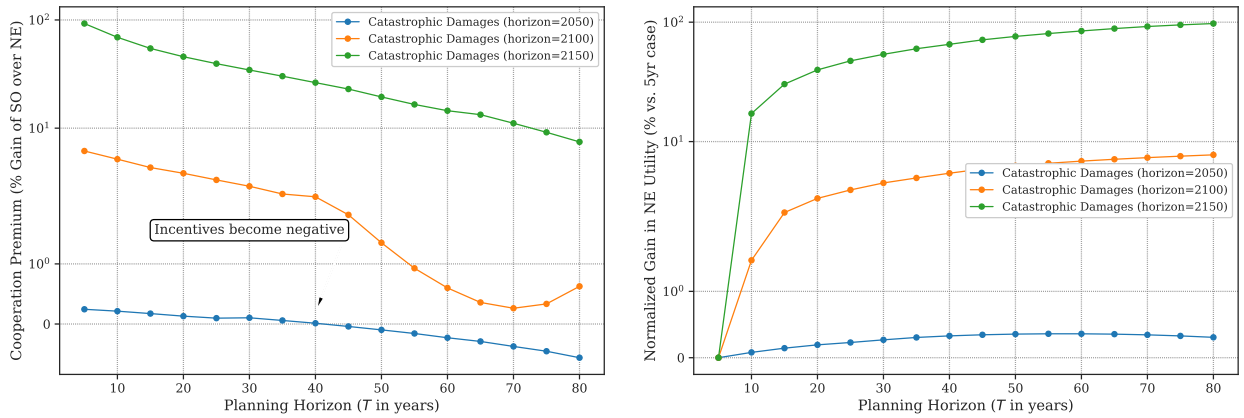
The *Cooperation Premium* for a specific terminal year is calculated as the relative welfare gain of the Social Optimum over the Nash Equilibrium, accumulated over that specific duration:

$$\text{Cooperation Premium}(T; T_{\text{end}}) = \frac{W^{\text{SO}}(T, T_{\text{end}}) - W^{\text{NE}}(T, T_{\text{end}})}{\left|W^{\text{NE}}(T, T_{\text{end}})\right|} \times 100. \tag{14}$$

Similarly, the *Value of Foresight* measures the private economic incentive to extend one's planning horizon within a game of fixed duration T_{end}. It is defined as the percentage improvement in non-cooperative welfare relative to the myopic baseline ($T = 1$, i.e., 5 years) calculated over the same duration:

$$\text{Value of Foresight}(T; T_{\text{end}}) = \frac{W^{\text{NE}}(T, T_{\text{end}}) - W^{\text{NE}}(1, T_{\text{end}})}{\left|W^{\text{NE}}(1, T_{\text{end}})\right|} \times 100. \tag{15}$$

Figure 5 presents a critical result. We replicate the analysis from Figure 4 for the *Catastrophic Damages world*, but for three different terminal game horizons: 2050, 2100, and 2150. The *y*-axis in both panels is on a logarithmic scale to capture the full range of outcomes.



(a) Cooperation Premium vs. Terminal Horizon

(b) Value of Foresight vs. Terminal Horizon

Figure 5: **Economic Incentives are Exponentially Sensitive to the Terminal Time Horizon.** Both panels plot outcomes in the *Catastrophic Damages world* for games ending in 2050, 2100, and 2150. Note the logarithmic scale on the *y*-axis. (a) The Cooperation Premium (percentage gain of SO over NE) is negligible for a 2050 horizon, significant for 2100, and an order of magnitude larger for 2150. (b) Similarly, the private economic value of long-term planning (normalized gain in NE utility) is close to zero for a 2050 horizon but becomes immense when the horizon is extended to 2150.

The findings demonstrate that the economic stakes of climate policy are critically dependent on the time horizon that policymakers are willing to consider.

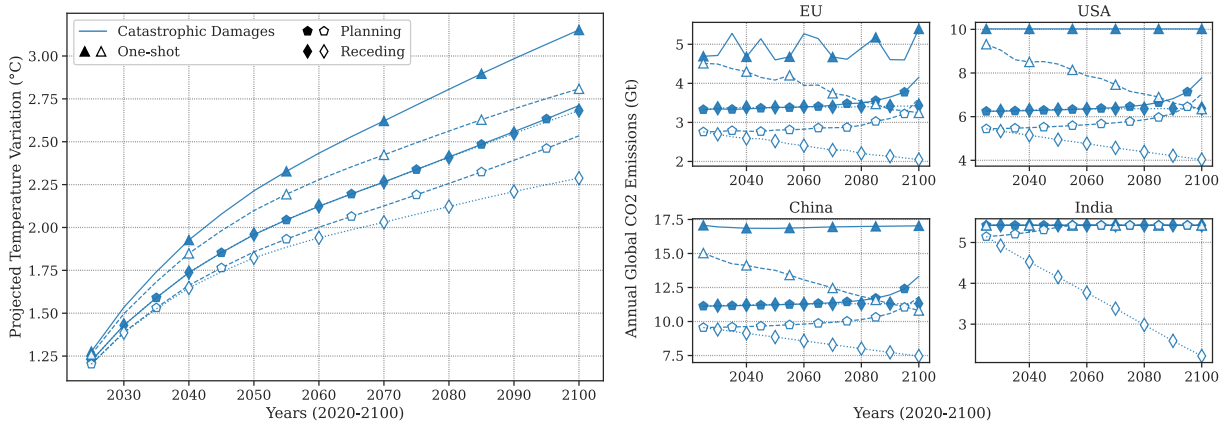
Panel 5a shows the impact on the incentive to cooperate. If the game is assumed to end in 2050 (blue line), the Cooperation Premium is near zero and even negative as the planning horizon increases: the shift occurs when the planning horizon is 2060 (even if the utility functions are truncated at 2050). In such a truncated world, the worst consequences of climate change are never realized, and the game strategically resembles the *Marginal Damages* case. Extending the horizon to 2100 (orange line) reveals the significant premium discussed in our previous section. However, extending the horizon further to 2150 (green line) causes the premium to increase by another order of magnitude, starting at nearly 100% (a doubling of welfare) for myopic players (5-year horizon).

Panel 5b reveals an even more concerning effect on the private incentive for long-term planning. The economic gains from foresight for a non-cooperative player are negligible with a 2050 horizon. With a 2100 horizon, they are significant. But with a 2150 horizon, they become astronomical. For example, a self-interested player adopting a 30-year planning horizon (relative to a 5-year one) sees their utility increase by nearly 50%. This demonstrates that unilateral, far-sighted climate action is not just an ethical choice but a natural, rational economic strategy, provided the policymaker's horizon is sufficiently long (75 years in this case).

This leads to our third major result: the standard practice of limiting climate-economic analysis at 2100 may cause a **systematic and severe underestimation of the true economic incentives for both international cooperation and unilateral long-term planning**. The "paradox of cooperation" is, in large part, an artifact of an artificially shortened future. When the long-run, path-dependent nature of catastrophic climate damage is properly accounted for, the economic case for ambitious climate policy becomes overwhelmingly strong.

6.5. The Game-Theoretic Superiority of the Receding Horizon Framework

Having established the critical importance of perceived economic damage and strategic foresight, we now address a central question for policy design: which institutional framework is best suited to navigate this high-stakes environment? We compare the three frameworks defined in our experimental design within the *Catastrophic Damages world*: myopic *One-stage*, full-commitment *Planning*, and adaptive *Receding Horizon*. Figure 6 presents the results, providing a clear hierarchy of policy effectiveness.



(a) Projected Global Temperature Variation

(b) Individual Emissions Pathways

Figure 6: **The Perils of Myopia and the Superiority of an Adaptive Policy Framework.** All simulations are conducted under the Catastrophic damage function. (**A**) Projected global temperature variation for non-cooperative (NE, filled markers) and socially optimal (SO, open markers) equilibria under the three policy concepts. (**B**) The corresponding individual emissions pathways for four key players/blocs. The disastrous temperature outcomes of the *One-shot* concept are driven by the myopic, high-emission strategies of players like the USA and India. The *Receding Horizon* framework consistently yields the best environmental outcomes.

The analysis reveals three major findings. First, the results show that myopic policymaking leads to **disastrous consequences**. The *One-shot* trajectories (triangles) lead to catastrophically high warming, with the non-cooperative *One-shot NE* exceeding 3.0°C. The Figure 6b panels reveal the mechanism: myopic players with high marginal benefits from emitting (e.g., USA, India) have a dominant strategy to emit at or near their maximum capacity, as they are blind to the devastating long-term consequences.

Second, the results confirm the **value of strategic foresight**. Both the *Planning* (pentagons) and *Receding Horizon* (diamonds) frameworks, which incorporate long-term consequences, lead to outcomes in a much safer climate space as illustrated in Figure 6a. The simple act of planning, even non-cooperatively, is sufficient to discipline players away from the most reckless emissions strategies and to avoid the worst of the climate cliff (the part where the damage function increases exponentially).

Third, and most critically for policy design, our findings reveal the **structural superiority of the Receding Horizon framework**. In both the non-cooperative and cooperative cases, the *Receding Horizon* concept (diamonds) yields a better end-of-century climate outcome than the static *Planning* concept as illustrated in Figure 6a. The fully cooperative *Receding SO* path (open diamonds, dotted line) results in the lowest final temperature of any scenario.

The mechanism behind this superiority is revealed in the emissions pathways shown in Figure 6b. The static *Planning* framework, while far-sighted, suffers from a critical flaw known as a *terminal date effect*. Because players optimize only up to the year 2100, they treat damages beyond that date as having zero cost. This creates a perverse incentive to defer abatement efforts toward the end of the planning period, as the long-run consequences of late-stage emissions are not fully internalized. This behavior is evident in India's *Planning SO* path, which is relatively flat compared to the steep and continuous emissions decline incentivized by the adaptive *Receding Horizon* framework.

The *Receding Horizon* framework solves this problem through its process of periodic re-optimization. By constantly extending the planning window into the future, it ensures that there is no final period where the long-term consequences of emissions can be ignored. This provides incentives for a steady and sustained abatement path from the outset, particularly for developing economies, and prevents the *end-of-game* strategic behavior that can undermine static long-term plans.

This provides our final major result: institutional designs that encourage or require iterative re-commitment and periodic updates to climate strategies are not merely politically pragmatic; they are game-theoretically superior for achieving more ambitious environmental goals. The underlying principle of periodically reviewing and strengthening commitments finds strong theoretical support in this strategic framework.

6.6. *From Aggregate Gains to Individual Incentives: A Pareto Analysis*

Our analysis has established that cooperative and far-sighted frameworks are superior in aggregate terms. However, the stability of any agreement hinges on individual player incentives. To better understand this, we must disaggregate the results and examine the trade-offs facing each player. Figure 7 provides this disaggregated view, plotting each player's total utility against the resulting 2100 temperature outcome for the six key solution concepts, all within the *Catastrophic damage* world.

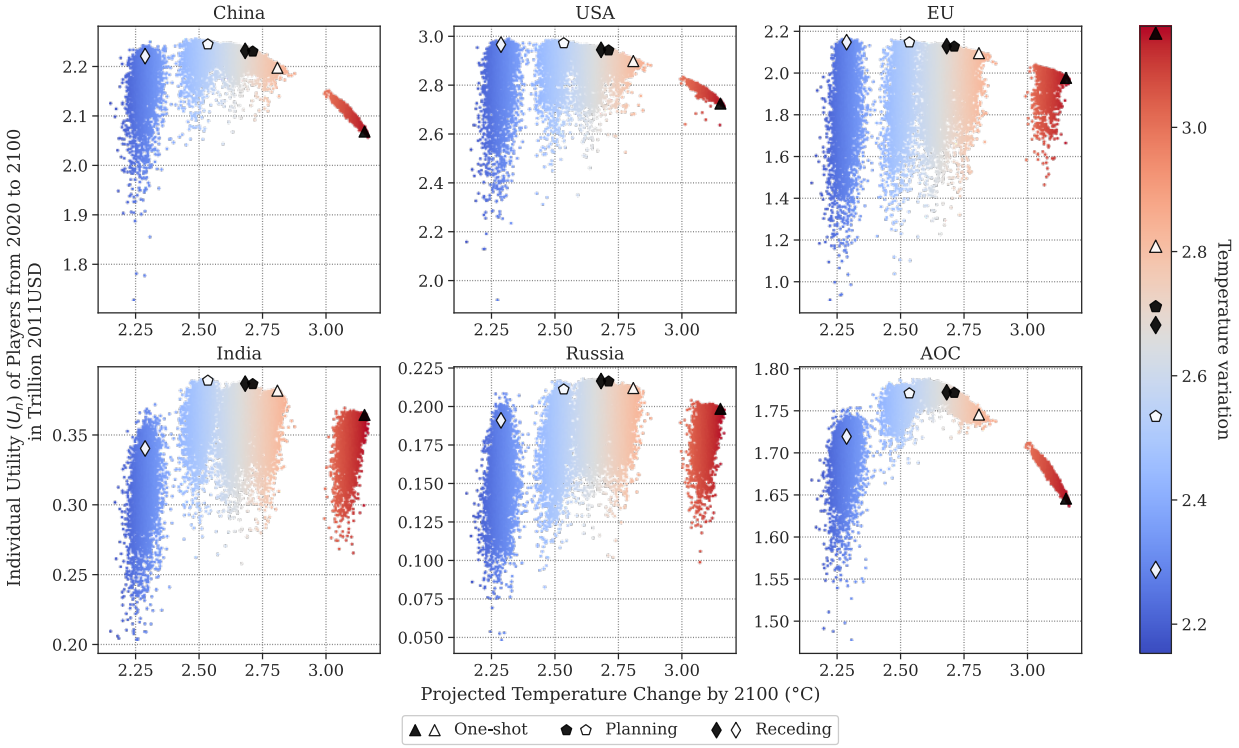


Figure 7: **Individual Player Trade-offs Between Utility and Climate Outcomes.** Each panel plots the total utility for a single player (y-axis) against the projected 2100 temperature (x-axis) under the *Catastrophic damage* scenario. The scatter plot shows the space of possible outcomes, with the color indicating the temperature. The six key equilibrium concepts are overlaid: NE (filled markers) and SO (open markers) for the *One-shot* (triangle), *Planning* (pentagon), and *Receding Horizon* (diamond) frameworks. The figure reveals the heterogeneous nature of the *Free-Rider Problem*, as the utility cost of achieving better climate outcomes varies significantly across players.

Each panel in the figure represents the strategic choice set for an individual player. The cloud of possible outcomes demonstrates the fundamental trade-off: achieving better environmental outcomes (moving left to lower temperatures) generally requires accepting a lower private utility. The equilibrium concepts represent key focal points within this trade-off space. This disaggregated view yields two critical insights.

First, the analysis confirms the universal irrationality of myopic policy. For every player, the *One-shot NE* outcome (solid black triangle) is in an unequivocally poor position: it results in the highest temperature and a suboptimal level of utility compared to other non-cooperative strategies. Every player is better off by incorporating some degree of foresight, even unilaterally.

Second, and more critically, the figure reveals the heterogeneous nature of the *Free-Rider Problem* by visualizing the individual *cost of cooperation*. This can be seen by comparing the non-cooperative and cooperative outcomes for the best-performing framework, the *Receding Horizon* (comparing the solid black diamond to the open white diamond).

- For players like **China, the USA, and the EU**, the cooperative outcome (Receding SO) is actually positioned higher or at the same level on the utility axis than the non-cooperative one (Receding NE) and performs better

than the other solution concept. For these developed or major economies, the avoided catastrophic damages are so valuable that they realize a net utility gain from full cooperation.

- For players like **India, Russia, and All-Other-Countries**, the story is starkly different. The *Receding SO* outcome results in a significantly lower utility than the *Receding NE* and other solution concepts. For these players, the strict abatement required by the Grand Coalition imposes a development cost that outweighs their specific avoided damages. Their incentive to free-ride is therefore maximizing; they are rationally better off in the non-cooperative planning equilibrium than in the social optimum. This result numerically confirms the structural tension identified in Corollary 1: the Grand Coalition is not internally stable because specific key players face a substantial private loss from joining.

Ultimately, this analysis provides a clear and sobering conclusion. While a shared perception of catastrophic damage creates a large collective incentive to form the Grand Coalition, it does not guarantee its stability. The cost of cooperation is not borne equally, and the *Free-Rider Incentive* persists. This confirms the foundational tension of IEA theory: the most effective agreement (the Grand Coalition) is not stable.

7. Conclusion

This paper bridges the gap between the strategic abstraction of IEA literature and the geophysical complexity of IAM. By embedding a nonlinear, path-dependent climate module directly into the utility structure of a dynamic open-loop game, we proved that the resulting interaction constitutes a state-transitive weighted potential game. This theoretical contribution renders a complex system analytically tractable, guaranteeing the existence of a pure-strategy open-loop Nash Equilibrium and providing sufficient conditions for its uniqueness. Our numerical analysis, calibrated to real-world economic and geophysical data, yields three distinct insights into the incentives for climate cooperation.

First, we show that the "paradox of cooperation" is endogenous to the damage function. In a "Marginal Damages" world, the free-rider incentive dominates, and the gains from cooperation are negligible. However, in a "Catastrophic Damage" world characterized by tipping points, the strategic landscape shifts. The threat of nonlinear damages tames non-cooperative players, significantly increasing the welfare gains from cooperation and making the formation of a Grand Coalition a first-order economic priority.

Second, we demonstrate the crucial role of the planning horizon. We find that limiting analysis to the year 2100 creates a "tyranny of the terminal horizon" that systematically underestimates the economic case for action. Extending the horizon reveals that long-term foresight is not merely fair for future generations but individually rational. Crucially, we find that under conditions of serious climatic damage, strategic foresight acts as a partial substitute for coordination: as players extend their planning horizons, the NE converges significantly toward the SO, reducing the cooperation premium.

Third, we establish the game-theoretic superiority of the adaptive framework. We compared a static "Full Commitment" approach against an adaptive "Receding Horizon" one. The adaptive approach, which mimics the periodic re-commitment structure of the Paris Agreement, consistently yields superior environmental outcomes. By continuously pushing the planning window forward, the Receding Horizon framework eliminates the "terminal date effect" where abatement incentives collapse at the end of a fixed plan.

Our analysis relies on an open-loop equilibrium concept, which assumes a capacity for commitment over the planning period. While this provides a necessary benchmark, future research should explore Feedback or Markov Perfect Equilibria to capture the purely dynamical nature of political decision-making. Nevertheless, our results confirm that when the physical reality of catastrophic damage is properly integrated into strategic models, the economic incentives for both unilateral foresight and international cooperation are far stronger than standard models suggest.

Appendix A. The Simple Climate Model (SCM) in Detail

Appendix A.1. State-Space Representation

As stated in the main text, the geophysical model links emissions E_t to the change in global mean atmospheric temperature θ^{at}_{t+1}. The system's state is $\boldsymbol{x}_t = (\boldsymbol{C}_t, \boldsymbol{\theta}_t)$. The dynamics are governed by two coupled linear state-space models, which we present here for completeness.

The carbon cycle dynamics are given by:

$$\begin{aligned} \boldsymbol{C}_{t+1} &= \boldsymbol{A}_C \boldsymbol{C}_t + \boldsymbol{b}_C E_t \\ C^{\text{at}}_{t+1} &= \boldsymbol{d}_C^\top \boldsymbol{C}_{t+1} \end{aligned} \tag{A.1}$$

where $\boldsymbol{C}_t \in \mathbb{R}^m$ is the vector of carbon stocks, $\boldsymbol{A}_C$ is the $m \times m$ transition matrix, $\boldsymbol{b}_C \in \mathbb{R}^m$ is the emissions distribution vector, and $\boldsymbol{d}_C \in \mathbb{R}^m$ is a selection vector that extracts the atmospheric stock.

The temperature dynamics are driven by radiative forcing, F_{t+1}:

$$\begin{aligned} \boldsymbol{\theta}_{t+1} &= \boldsymbol{A}_\theta \boldsymbol{\theta}_t + \boldsymbol{b}_\theta F_{t+1} \\ \theta^{\text{at}}_{t+1} &= \boldsymbol{d}_\theta^\top \boldsymbol{\theta}_{t+1} \end{aligned} \tag{A.2}$$

where $\boldsymbol{\theta}_t \in \mathbb{R}^{m'}$ is the vector of thermal reservoir temperatures, $\boldsymbol{A}_\theta$ is the $m' \times m'$ transition matrix, $\boldsymbol{b}_\theta \in \mathbb{R}^{m'}$ maps forcing into the thermal reservoirs, and $\boldsymbol{d}_\theta \in \mathbb{R}^{m'}$ selects the atmospheric temperature. Radiative forcing, F_{t+1}, is given by:

$$F_{t+1} = F_{2\times\text{CO}_2} \log_2\left(\frac{C^{\text{at}}_{t+1}}{C_{\text{ref}}}\right) + F^{\text{exo}}_{t+1}, \tag{A.3}$$

where F^{exo}_{t+1} is exogenous emission that are not decided by players taken as in REMIND model (see [74]).

Appendix A.2. Derivation of the Reduced-Form State Transition Functions

For use in the main text, we derive the one-step relationship between aggregate emissions E_t and the resulting atmospheric carbon C^{at}_{t+1} and temperature θ^{at}_{t+1}.

Atmospheric Carbon Concentration. Substituting the first line of (A.1) into the second yields:

$$C^{\text{at}}_{t+1} = \boldsymbol{d}_C^\top(\boldsymbol{A}_C \boldsymbol{C}_t + \boldsymbol{b}_C E_t) = (\boldsymbol{d}_C^\top \boldsymbol{A}_C \boldsymbol{C}_t) + (\boldsymbol{d}_C^\top \boldsymbol{b}_C) E_t.$$

This corresponds to the reduced-form equation in the main text, $C^{\text{at}}_{t+1} = \psi_C(\boldsymbol{C}_t) + \beta_{\text{C}} E_t$, where we define:

- $\psi_C(\boldsymbol{C}_t) \equiv \boldsymbol{d}_C^\top \boldsymbol{A}_C \boldsymbol{C}_t$, representing the contribution from the prior carbon state.
- $\beta_{\text{C}} \equiv \boldsymbol{d}_C^\top \boldsymbol{b}_C$, representing the marginal contribution of current emissions.

Atmospheric Temperature. Substituting (A.3) into the first line of (A.2) and then substituting the result into the second line gives:

$$\begin{aligned} \theta^{\text{at}}_{t+1} &= \boldsymbol{d}_\theta^\top \left(\boldsymbol{A}_\theta \boldsymbol{\theta}_t + \boldsymbol{b}_\theta \left[F_{2\times\text{CO}_2} \log_2\left(\frac{C^{\text{at}}_{t+1}}{C_{\text{ref}}}\right) + F^{\text{exo}}_{t+1} \right] \right) \\ &= (\boldsymbol{d}_\theta^\top \boldsymbol{A}_\theta \boldsymbol{\theta}_t + \boldsymbol{d}_\theta^\top \boldsymbol{b}_\theta F^{\text{exo}}_{t+1}) + (\boldsymbol{d}_\theta^\top \boldsymbol{b}_\theta F_{2\times\text{CO}_2}) \log_2(C^{\text{at}}_{t+1}) - (\boldsymbol{d}_\theta^\top \boldsymbol{b}_\theta F_{2\times\text{CO}_2}) \log_2(C_{\text{ref}}) \end{aligned}$$

Using the identity $\log_2(x) = \ln(x)/\ln(2)$ and grouping terms, we can write:

$$\theta^{\text{at}}_{t+1} = \left[\boldsymbol{d}_\theta^\top \boldsymbol{A}_\theta \boldsymbol{\theta}_t + \boldsymbol{d}_\theta^\top \boldsymbol{b}_\theta F^{\text{exo}}_{t+1} - \frac{\boldsymbol{d}_\theta^\top \boldsymbol{b}_\theta F_{2\times\text{CO}_2}}{\ln(2)} \ln(C_{\text{ref}}) \right] + \left[\frac{\boldsymbol{d}_\theta^\top \boldsymbol{b}_\theta F_{2\times\text{CO}_2}}{\ln(2)} \right] \ln(C^{\text{at}}_{t+1}).$$

This corresponds to the reduced-form equation in the main text, $\theta^{\text{at}}_{t+1} = \psi_\theta(\boldsymbol{x}_t) + \beta_\theta \ln(C^{\text{at}}_{t+1})$, where we define:

- $\psi_\theta(\boldsymbol{x}_t) \equiv \boldsymbol{d}_\theta^\top \boldsymbol{A}_\theta \boldsymbol{\theta}_t + \boldsymbol{d}_\theta^\top \boldsymbol{b}_\theta F^{\text{exo}}_{t+1} - \beta_\theta \ln(C_{\text{ref}})$, representing the temperature evolution based on the prior thermal state and exogenous forcing.
- $\beta_\theta \equiv (\boldsymbol{d}_\theta^\top \boldsymbol{b}_\theta F_{2\times\text{CO}_2})/\ln(2)$, representing the temperature sensitivity to atmospheric carbon concentration.

Table A.2: Key Geophysical Model Parameters

Parameter	Symbol	Value	Unit
Pre-industrial CO_2 concentration	C_{ref}	592.14	GtC
Forcing from CO_2 doubling	$F_{2\times CO_2}$	3.7	W/m^2
Carbon Cycle Model	-	Joos et al. (2013)	-
Temperature Module	-	Geoffroy et al. (2013)	-

Appendix A.3. SCM Parameterization

The numerical results presented in this paper are based on standard, well-vetted SCMs from the literature. Specifically, the carbon cycle model is parameterized according to Joos et al. (2013), and the temperature module follows Geoffroy et al. (2013). The specific values for the matrices and vectors are available in the authors' publicly available replication code. Key global parameters are listed in Table A.2.

Appendix B. Proofs for Section 4

Appendix B.1. Proof of Proposition 6: General Equilibrium KKT Conditions

Proof. By the definition of a Nash Equilibrium (Definition 2), the strategy vector $\boldsymbol{a}_n^*$ for any player n must be an optimal solution to that player's individual utility maximization problem, holding the strategies of all other players $\boldsymbol{a}_{-n}^*$ fixed. For each player n, this problem is:

$$\begin{aligned} \underset{\boldsymbol{a}_n \in \mathcal{A}_n^T}{\text{maximize}} \quad & U_n(\boldsymbol{a}_n, \boldsymbol{a}_{-n}^*) = \sum_{t=1}^{T} \left[B_n(a_{nt}) - w_n D(\theta_{t+1}^{\text{at}}(\boldsymbol{a}_n, \boldsymbol{a}_{-n}^*)) \right] \\ \text{subject to} \quad & a_{nt} - e_n^{\max} \leq 0, \quad \forall t \in \mathcal{T} \\ & e_n^{\min} - a_{nt} \leq 0, \quad \forall t \in \mathcal{T}. \end{aligned}$$

Under Assumptions 1 and 2, the objective function is continuously differentiable, and the constraints define a convex set. Therefore, the Karush-Kuhn-Tucker (KKT) conditions are necessary for optimality. We formulate the Lagrangian for player n's problem with non-negative multipliers $\overline{\lambda}_{nt}$ and $\underline{\lambda}_{nt}$ for the upper and lower bound constraints, respectively:

$$\mathcal{L}_n(\boldsymbol{a}_n, \overline{\boldsymbol{\lambda}}_n, \underline{\boldsymbol{\lambda}}_n) := U_n(\boldsymbol{a}_n, \boldsymbol{a}_{-n}^*) + \sum_{t=1}^{T} \overline{\lambda}_{nt}(e_n^{\max} - a_{nt}) + \sum_{t=1}^{T} \underline{\lambda}_{nt}(a_{nt} - e_n^{\min}).$$

The KKT conditions require that at the optimum $\boldsymbol{a}_n^*$, the partial derivative of the Lagrangian with respect to each choice variable a_{nt} must be zero (stationarity), and the complementary slackness conditions must hold.

The stationarity condition for each $t \in \mathcal{T}$ is:

$$\left.\frac{\partial \mathcal{L}_n}{\partial a_{nt}}\right|_{\boldsymbol{a}^*} = \left.\frac{\partial U_n}{\partial a_{nt}}\right|_{\boldsymbol{a}^*} - \overline{\lambda}_{nt} + \underline{\lambda}_{nt} = 0.$$

We now evaluate the partial derivative of the utility function. An emission in period t, a_{nt}, only affects the benefit function in period t and the damage function in periods $s \geq t$. Thus,

$$\left.\frac{\partial U_n}{\partial a_{nt}}\right|_{\boldsymbol{a}^*} = B_n'(a_{nt}^*) - w_n \sum_{s=t}^{T} \frac{\partial D(\theta_{s+1}^{\text{at}*})}{\partial a_{nt}}.$$

Let $\mathcal{M}_{nt}(\boldsymbol{a}^*)$ denote the private marginal cost of emissions for player n at time t, defined as the summation term:

$$\mathcal{M}_{nt}(\boldsymbol{a}^*) := \sum_{s=t}^{T} \frac{\partial D(\theta_{s+1}^{\text{at}*})}{\partial a_{nt}}.$$

Substituting this into the stationarity equation yields condition (i) of the proposition:

$$B_n'(a_{nt}^*) - w_n \mathcal{M}_{nt}(\boldsymbol{a}^*) + \underline{\lambda}_{nt} - \overline{\lambda}_{nt} = 0.$$

The complementary slackness conditions are standard and yield condition (ii):

$$\underline{\lambda}_{nt}(a_{nt}^* - e_n^{\min}) = 0 \quad \text{and} \quad \overline{\lambda}_{nt}(e_n^{\max} - a_{nt}^*) = 0.$$

We derive the expression for $\mathcal{M}_{nt}(\boldsymbol{a}^*)$ by applying the multivariable chain rule through the geophysical model. The derivative of damages in a future period $r \geq t$ with respect to emissions at t is:

$$\frac{\partial D(\theta_{r+1}^{\text{at}*})}{\partial a_{nt}} = D'(\theta_{r+1}^{\text{at}*}) \frac{\partial \theta_{r+1}^{\text{at}*}}{\partial a_{nt}}.$$

The term $\partial \theta_{r+1}^{\text{at}*} / \partial a_{nt}$ represents the marginal effect of an emission pulse at time t on the atmospheric temperature at a future time r. This temperature is affected by the entire history of forcing from t to r. The derivative is therefore a sum over this history:

$$\frac{\partial \theta_{r+1}^{\text{at}*}}{\partial a_{nt}} = \sum_{s=t}^{r} \frac{\partial \theta_{r+1}^{\text{at}*}}{\partial F_{s+1}} \frac{\partial F_{s+1}}{\partial C_{s+1}^{\text{at}*}} \frac{\partial C_{s+1}^{\text{at}*}}{\partial a_{nt}}.$$

The components of this derivative are the impulse response functions of the linear geophysical model and the derivative of the forcing equation:

- The effect of an emission pulse at t on atmospheric carbon at $s \geq t$: $\partial C_{s+1}^{\text{at}*} / \partial a_{nt} = \boldsymbol{d}_C^\top \boldsymbol{A}_C^{s-t} \boldsymbol{b}_C$.
- The effect of atmospheric carbon on forcing at s: $\partial F_{s+1} / \partial C_{s+1}^{\text{at}*} = F_{2\times\text{CO}_2} / \ln(2) C_{s+1}^{\text{at}*}$.
- The effect of a forcing pulse at s on atmospheric temperature at $r \geq s$: $\partial \theta_{r+1}^{\text{at}*} / \partial F_{s+1} = \boldsymbol{d}_\theta^\top \boldsymbol{A}_\theta^{r-s} \boldsymbol{b}_\theta$.

Substituting these into the expression for $\mathcal{M}_{nt}(\boldsymbol{a}^*) = \sum_{r=t}^{T} \partial D(\theta_{r+1}^{\text{at}*}) / \partial a_{nt}$ gives:

$$\mathcal{M}_{nt}(\boldsymbol{a}^*) = \sum_{r=t}^{T} D'(\theta_{r+1}^{\text{at}*}) \left[\sum_{s=t}^{r} \left(\boldsymbol{d}_\theta^\top \boldsymbol{A}_\theta^{r-s} \boldsymbol{b}_\theta \right) \left(\frac{F_{2\times\text{CO}_2}}{\ln(2) C_{s+1}^{\text{at}*}} \right) \left(\boldsymbol{d}_C^\top \boldsymbol{A}_C^{s-t} \boldsymbol{b}_C \right) \right].$$

The double summation is over the domain $\{(s, r) \mid t \leq r \leq T, t \leq s \leq r\}$. Changing the order of summation to the equivalent domain $\{(s, r) \mid t \leq s \leq T, s \leq r \leq T\}$ and rearranging terms yields the expression in the proposition:

$$\mathcal{M}_{nt}(\boldsymbol{a}^*) = \sum_{s=t}^{T} \sum_{r=s}^{T} \left(\frac{D'(\theta_{r+1}^{\text{at}*})}{\ln(2)} \cdot \frac{\boldsymbol{d}_\theta^\top \boldsymbol{A}_\theta^{r-s} \boldsymbol{b}_\theta F_{2\times\text{CO}_2}}{C_{s+1}^{\text{at}*}} \cdot \boldsymbol{d}_C^\top \boldsymbol{A}_C^{s-t} \boldsymbol{b}_C \right).$$

Because these necessary conditions must hold for every player $n \in \mathcal{N}$ and every time period $t \in \mathcal{T}$, the proof is complete. □

Appendix B.2. Proof of Proposition 7: Equilibrium Ordering

Proof. The proof proceeds by contradiction. Assume that for some period t, the equilibrium ordering does not hold. This implies there exist two players, n and k, such that their emissions are not co-ordered across the two equilibria, i.e., $a_{nt} < \tilde{a}_{nt}$ and $a_{kt} > \tilde{a}_{kt}$.

From Proposition 6, any Nash Equilibrium must satisfy the KKT conditions. For any player $j \in \mathcal{N}$ and any period t, the stationarity condition is:

$$B_{jt}'(a_{jt}) - w_j \mathcal{M}_{jt}(\boldsymbol{a}) + \underline{\lambda}_{jt} - \overline{\lambda}_{jt} = 0.$$

A crucial observation is that the private marginal cost term, $\mathcal{M}_{jt}(\boldsymbol{a})$, is independent of the player index j because emissions are a uniformly mixed pollutant (as noted in Section 2). We can therefore denote this common marginal cost at equilibrium $\boldsymbol{a}$ by $\mathcal{M}_t(\boldsymbol{a})$. Rearranging the stationarity condition, we have for any player j:

$$\mathcal{M}_t(\boldsymbol{a}) = \frac{B_j'(a_{jt}) + \underline{\lambda}_{jt} - \overline{\lambda}_{jt}}{w_j}.$$

Since this must hold for both players n and k in both equilibria $\boldsymbol{a}$ and $\tilde{\boldsymbol{a}}$, we can write:

$$\frac{B_n'(a_{nt}) + \underline{\lambda}_{nt} - \overline{\lambda}_{nt}}{w_n} = \frac{B_k'(a_{kt}) + \underline{\lambda}_{kt} - \overline{\lambda}_{kt}}{w_k} \tag{B.1}$$

$$\frac{B_n'(\tilde{a}_{nt}) + \underline{\tilde{\lambda}}_{nt} - \overline{\tilde{\lambda}}_{nt}}{w_n} = \frac{B_k'(\tilde{a}_{kt}) + \underline{\tilde{\lambda}}_{kt} - \overline{\tilde{\lambda}}_{kt}}{w_k} \tag{B.2}$$

Our initial assumption about the ordering of actions allows us to simplify the Lagrange multipliers via the complementary slackness conditions.

- The condition $a_{nt} < \tilde{a}_{nt}$ implies that a_{nt} cannot be at the upper bound ($a_{nt} < e_n^{\max}$) and $\tilde{a}_{nt}$ cannot be at the lower bound ($\tilde{a}_{nt} > e_n^{\min}$). This requires $\overline{\lambda}_{nt} = 0$ and $\underline{\tilde{\lambda}}_{nt} = 0$.
- Similarly, the condition $a_{kt} > \tilde{a}_{kt}$ implies that a_{kt} cannot be at the lower bound ($a_{kt} > e_k^{\min}$) and $\tilde{a}_{kt}$ cannot be at the upper bound ($\tilde{a}_{kt} < e_k^{\max}$). This requires $\underline{\lambda}_{kt} = 0$ and $\overline{\tilde{\lambda}}_{kt} = 0$.

Substituting these into Equations (B.1) and (B.2) yields:

$$\frac{B_n'(a_{nt}) + \underline{\lambda}_{nt}}{w_n} = \frac{B_k'(a_{kt}) - \overline{\lambda}_{kt}}{w_k} \quad \text{and} \quad \frac{B_n'(\tilde{a}_{nt}) - \overline{\tilde{\lambda}}_{nt}}{w_n} = \frac{B_k'(\tilde{a}_{kt}) + \underline{\tilde{\lambda}}_{kt}}{w_k}$$

Subtracting the second identity from the first gives:

$$\frac{B_n'(a_{nt}) - B_n'(\tilde{a}_{nt}) + \underline{\lambda}_{nt} + \overline{\tilde{\lambda}}_{nt}}{w_n} = \frac{B_k'(a_{kt}) - B_k'(\tilde{a}_{kt}) - \overline{\lambda}_{kt} - \underline{\tilde{\lambda}}_{kt}}{w_k}.$$

We now analyze the sign of the numerators.

- On the left-hand side: Since B_n is strictly concave, its derivative B_n' is strictly decreasing. Our assumption $a_{nt} < \tilde{a}_{nt}$ thus implies $B_n'(a_{nt}) > B_n'(\tilde{a}_{nt})$. As the Lagrange multipliers are non-negative ($\underline{\lambda}_{nt} \geq 0, \overline{\tilde{\lambda}}_{nt} \geq 0$), the entire numerator is strictly positive.
- On the right-hand side: Similarly, since B_k is strictly concave, B_k' is strictly decreasing. Our assumption $a_{kt} > \tilde{a}_{kt}$ implies $B_k'(a_{kt}) < B_k'(\tilde{a}_{kt})$. As the Lagrange multipliers are non-negative, the terms $-\overline{\lambda}_{kt}$ and $-\underline{\tilde{\lambda}}_{kt}$ are non-positive. Therefore, the entire numerator is strictly negative.

The equality requires a strictly positive number to be equal to a strictly negative number, which is a contradiction. Thus, our initial assumption must be false. It must be that if $a_{nt} < \tilde{a}_{nt}$ for some player n, then $a_{kt} \leq \tilde{a}_{kt}$ for all $k \in \mathcal{N}$.

To prove the final statement on boundary conditions, assume $a_{kt} = \tilde{a}_{kt}$ for some player k, where a_{kt} is an interior point: $a_{kt} \in (e_k^{\min}, e_k^{\max})$. By complementary slackness, all of player k's Lagrange multipliers must be zero: $\underline{\lambda}_{kt} = \overline{\lambda}_{kt} = \underline{\tilde{\lambda}}_{kt} = \overline{\tilde{\lambda}}_{kt} = 0$. The two KKT equalities become:

$$\frac{B_n'(a_{nt}) + \underline{\lambda}_{nt}}{w_n} = \frac{B_k'(a_{kt})}{w_k} \quad \text{and} \quad \frac{B_n'(\tilde{a}_{nt}) - \overline{\tilde{\lambda}}_{nt}}{w_n} = \frac{B_k'(\tilde{a}_{kt})}{w_k}$$

Since $a_{kt} = \tilde{a}_{kt}$, the right-hand sides are identical. Equating the left-hand sides and rearranging gives:

$$B_n'(a_{nt}) - B_n'(\tilde{a}_{nt}) = -(\underline{\lambda}_{nt} + \overline{\tilde{\lambda}}_{nt}).$$

The left-hand side is strictly positive (as established before), while the right-hand side is non-positive. This is a contradiction. Therefore, if $a_{kt} = \tilde{a}_{kt}$, the action must lie on a boundary, $a_{kt} \in \{e_k^{\min}, e_k^{\max}\}$. □

Appendix B.3. Proof of Proposition 8 (Sufficient Conditions for a Unique NE)

The proposition states three independent sufficient conditions for the uniqueness of the Nash Equilibrium in the one-stage game Γ_T. We prove that if any of these conditions hold, the existence of two distinct equilibria leads to a contradiction.

Let $\boldsymbol{a}^*$ and $\tilde{\boldsymbol{a}}^*$ be two distinct pure-strategy Nash Equilibria. Since they are distinct, there must be a first time period, let's call it $t_0 \in \{1, \dots, T\}$, where their action profiles differ. By definition, $\boldsymbol{a}_t^* = \tilde{\boldsymbol{a}}_t^*$ for all $t < t_0$, and $\boldsymbol{a}_{t_0}^* \neq \tilde{\boldsymbol{a}}_{t_0}^*$. This implies that the state of the system is identical for both trajectories at the beginning of period t_0, i.e., $\boldsymbol{x}_{t_0} = \tilde{\boldsymbol{x}}_{t_0}$. Without loss of generality, we can assume that $t_0 = 1$ by setting $\boldsymbol{x}_1 = \boldsymbol{x}_{t_0}$.

By Proposition 7 (Equilibrium Ordering), this implies that for all players $k \in \mathcal{N}$, $a_k^* \leq \tilde{a}_k^*$. Consequently, the aggregate emissions are ordered: $E^* = \sum_k a_k^* < \sum_k \tilde{a}_k^* = \tilde{E}^*$. This leads to ordered atmospheric carbon concentrations, $C^{\text{at}*} < \tilde{C}^{\text{at}*}$, and temperatures, $\theta^{\text{at}*} < \tilde{\theta}^{\text{at}*}$.

Both equilibria must satisfy the KKT conditions from Proposition 6. For player n, since $a_n^* < \tilde{a}_n^*$, the complementary slackness conditions imply that $\overline{\lambda}_n^* = 0$ (as a_n^* is not at the upper bound) and $\underline{\tilde{\lambda}}_n^* = 0$ (as $\tilde{a}_n^*$ is not at the lower bound). The first-order conditions for player n for each equilibrium are therefore:

$$B_n'(a_n^*) + \underline{\lambda}_n^* = w_n \mathcal{M}_n(\boldsymbol{a}^*) \tag{B.3}$$

$$B_n'(\tilde{a}_n^*) - \overline{\tilde{\lambda}}_n^* = w_n \mathcal{M}_n(\tilde{\boldsymbol{a}}^*) \tag{B.4}$$

where for the one-stage game, the private marginal cost is $\mathcal{M}_n(\boldsymbol{a}) = D'(\theta^{\text{at}}(\boldsymbol{a})) \cdot \beta_\theta b_c' / C^{\text{at}}(\boldsymbol{a})$.

Proof of (i): Rapidly Accelerating Damages. Subtracting (B.4) from (B.3) and rearranging yields:

$$\frac{B_n'(a_n^*) - B_n'(\tilde{a}_n^*) + \underline{\lambda}_n^* + \overline{\tilde{\lambda}}_n^*}{w_n \beta_\theta b_c'} = \frac{D'(\theta^{\text{at}*})}{C^{\text{at}*}} - \frac{D'(\tilde{\theta}^{\text{at}*})}{\tilde{C}^{\text{at}*}}.$$

By Assumption 1, B_n is strictly concave, so B_n' is strictly decreasing. Since $a_n^* < \tilde{a}_n^*$, we have $B_n'(a_n^*) - B_n'(\tilde{a}_n^*) > 0$. As the Lagrange multipliers are non-negative, the left-hand side of the equation is strictly positive. This implies:

$$0 < \frac{D'(\theta^{\text{at}*})}{C^{\text{at}*}} - \frac{D'(\tilde{\theta}^{\text{at}*})}{\tilde{C}^{\text{at}*}} \implies \frac{D'(\tilde{\theta}^{\text{at}*})}{\tilde{C}^{\text{at}*}} < \frac{D'(\theta^{\text{at}*})}{C^{\text{at}*}} \implies \frac{D'(\tilde{\theta}^{\text{at}*})}{D'(\theta^{\text{at}*})} < \frac{\tilde{C}^{\text{at}*}}{C^{\text{at}*}}.$$

This contradicts condition (i) of the proposition, which states $D'(\tilde{\theta}^{\text{at}*})/D'(\theta^{\text{at}*}) \geq \tilde{C}^{\text{at}*}/C^{\text{at}*}$. Thus, two such distinct equilibria cannot exist.

Proof of (ii): Rapidly Diminishing Benefits. Rearrange the KKT conditions in (B.3) and (B.4) and multiply by their respective carbon concentrations:

$$B_n'(a_n^*) C^{\text{at}*} = w_n \beta_\theta b_c' D'(\theta^{\text{at}*}) - \underline{\lambda}_n^* C^{\text{at}*} \quad \text{and} \quad B_n'(\tilde{a}_n^*) \tilde{C}^{\text{at}*} = w_n \beta_\theta b_c' D'(\tilde{\theta}^{\text{at}*}) + \overline{\tilde{\lambda}}_n^* \tilde{C}^{\text{at}*}.$$

Subtracting the first equation from the second gives:

$$B_n'(\tilde{a}_n^*) \tilde{C}^{\text{at}*} - B_n'(a_n^*) C^{\text{at}*} = w_n \beta_\theta b_c' (D'(\tilde{\theta}^{\text{at}*}) - D'(\theta^{\text{at}*})) + (\overline{\tilde{\lambda}}_n^* \tilde{C}^{\text{at}*} + \underline{\lambda}_n^* C^{\text{at}*}).$$

By Assumption 2, D is strictly convex, so D' is strictly increasing. Since $\theta^{\text{at}*} < \tilde{\theta}^{\text{at}*}$, we have $D'(\tilde{\theta}^{\text{at}*}) - D'(\theta^{\text{at}*}) > 0$. The term involving the Lagrange multipliers is also strictly positive (as at least one of a_n^* or $\tilde{a}_n^*$ must be interior for the two equilibria to be distinct). Therefore, the right-hand side is strictly positive. This implies:

$$0 < B_n'(\tilde{a}_n^*) \tilde{C}^{\text{at}*} - B_n'(a_n^*) C^{\text{at}*} \implies B_n'(a_n^*) C^{\text{at}*} < B_n'(\tilde{a}_n^*) \tilde{C}^{\text{at}*} \implies \frac{C^{\text{at}*}}{\tilde{C}^{\text{at}*}} < \frac{B_n'(\tilde{a}_n^*)}{B_n'(a_n^*)}.$$

This contradicts condition (ii) of the proposition, which states $B_n'(\tilde{a}_n^*)/B_n'(a_n^*) \leq C^{\text{at}*}/C^{\text{at}*}$. Thus, two such distinct equilibria cannot exist.

Proof of (iii): Sufficiently Convex Damage Function. The existence of a Nash Equilibrium is equivalent to the existence of a maximizer of the potential function $\phi(\boldsymbol{a})$ from (5). A unique Nash Equilibrium is guaranteed if $\phi(\boldsymbol{a})$ has a unique maximizer, which is true if $\phi(\boldsymbol{a})$ is strictly concave in $\boldsymbol{a}$. The potential function is:

$$\phi(\boldsymbol{a}) = \sum_{k=1}^{N} \frac{B_k(a_k)}{w_k} - D(\theta^{\text{at}}(\boldsymbol{a})).$$

The first term, $\sum(B_k/w_k)$, is strictly concave in $\boldsymbol{a}$ because each B_k is strictly concave by Assumption 1. Thus, $\phi(\boldsymbol{a})$ is strictly concave if the second term, $-D(\theta^{\text{at}}(\boldsymbol{a}))$, is concave in $\boldsymbol{a}$. This requires the Hessian of $-D(\theta^{\text{at}}(\boldsymbol{a}))$ with respect to $\boldsymbol{a}$ to be negative semi-definite. The (i, j)-th element of this Hessian is $-\partial^2 D(\theta^{\text{at}})/\partial a_i \partial a_j$. All these partial derivatives are equal. Let us compute the second derivative with respect to a single action a_n:

$$\begin{aligned}
\frac{\partial D(\theta^{\text{at}})}{\partial a_n} &= D'(\theta^{\text{at}}) \frac{\partial \theta^{\text{at}}}{\partial a_n} = D'(\theta^{\text{at}}) \frac{\beta_\theta}{C^{\text{at}}} \beta_{\text{C}}. \\
\frac{\partial^2 D(\theta^{\text{at}})}{\partial a_n^2} &= \beta_{\text{C}} \frac{\partial}{\partial a_n} \left(\frac{D'(\theta^{\text{at}}) \beta_\theta}{C^{\text{at}}} \right) = \beta_{\text{C}} \beta_\theta \left[\frac{C^{\text{at}} \cdot \partial D'(\theta^{\text{at}})/\partial a_n - D'(\theta^{\text{at}}) \cdot \partial C^{\text{at}}/\partial a_n}{(C^{\text{at}})^2} \right] \\
&= \frac{(\beta_{\text{C}})^2 \beta_\theta}{(C^{\text{at}})^2} \left[C^{\text{at}} \cdot D''(\theta^{\text{at}}) \frac{\beta_\theta}{C^{\text{at}}} - D'(\theta^{\text{at}}) \right] = \frac{(\beta_{\text{C}})^2}{(C^{\text{at}})^2} \left[D''(\theta^{\text{at}}) \beta_\theta - D'(\theta^{\text{at}}) \right].
\end{aligned}$$

For $-D(\theta^{\text{at}})$ to be concave, we require $\partial^2(-D)/\partial a_n^2 \leq 0$, which is equivalent to $\partial^2 D/\partial a_n^2 \geq 0$. This condition holds if $D''(\theta^{\text{at}})\beta_\theta - D'(\theta^{\text{at}}) \geq 0$, which is precisely condition (iii). Since the Hessian matrix of $-D$ has this term on its diagonal and off-diagonal, this condition is sufficient for its negative semi-definiteness, ensuring concavity. As the first part of ϕ is strictly concave, the entire potential function is strictly concave, guaranteeing a unique maximizer and thus a unique Nash Equilibrium. □

Appendix B.4. Proof of Proposition 9

The proposition states two independent sufficient conditions for the uniqueness of the Nash Equilibrium in the planning game. We prove each condition separately.

Proof of (i): Strategic Substitutes. The proof proceeds in two main parts. First, we establish that the stated condition on $g_s(\boldsymbol{E})$ ensures the game is submodular (i.e., has strategic substitutes). Second, we use this structural property to prove that the equilibrium must be unique.

Establishing that the Game has Strategic Substitutes. A game has strategic substitutes if the marginal utility of any player's action is non-increasing in the actions of other players (or other actions of the same player at different times). This is true if the cross-partial derivatives of the utility function are non-positive: $\partial^2 U_n/\partial a_{nt}\partial a_{ks} \leq 0$ for all $(n, t) \neq (k, s)$.

The utility for player n is $U_n(\boldsymbol{a}) = \sum_{\tau=1}^{T} B_n(a_{n\tau}) - w_n \sum_{\tau=1}^{T} D(\theta^{\text{at}}_{\tau+1})$. The partial derivative with respect to a_{nt} is:

$$\frac{\partial U_n}{\partial a_{nt}} = B_n'(a_{nt}) - w_n \mathcal{M}_{nt}(\boldsymbol{a}),$$

where $\mathcal{M}_{nt}(\boldsymbol{a})$ is the private marginal cost. Differentiating with respect to another action, a_{ks}, gives:

$$\frac{\partial^2 U_n}{\partial a_{nt}\partial a_{ks}} = -w_n \frac{\partial \mathcal{M}_{nt}(\boldsymbol{a})}{\partial a_{ks}}.$$

Since $w_n > 0$, the condition for strategic substitutes ($\partial^2 U_n/\partial a_{nt}\partial a_{ks} \leq 0$) is equivalent to requiring $\partial \mathcal{M}_{nt}(\boldsymbol{a})/\partial a_{ks} \geq 0$.

The marginal cost is $\mathcal{M}_{nt}(\boldsymbol{a}) = \sum_{r=t}^{T} K_{r,t} \cdot g_r(\boldsymbol{E}(\boldsymbol{a}))$, where $K_{r,t}$ are positive constants from the geophysical model. Its derivative with respect to a_{ks} is:

$$\frac{\partial \mathcal{M}_{nt}}{\partial a_{ks}} = \sum_{r=\max(t,s)}^{T} K_{r,t} \cdot \frac{\partial g_r(\boldsymbol{E})}{\partial E_s} \cdot \frac{\partial E_s}{\partial a_{ks}} = \sum_{r=\max(t,s)}^{T} K_{r,t} \cdot \frac{\partial g_r(\boldsymbol{E})}{\partial E_s}.$$

The proposition's condition is that $g_r(\boldsymbol{E})$ is non-decreasing in emissions E_s, which means $\partial g_r(\boldsymbol{E})/\partial E_s \geq 0$. Since $K_{r,t} > 0$, every term in the summation is non-negative. Therefore, $\partial \mathcal{M}_{nt}/\partial a_{ks} \geq 0$, which confirms the game has strategic substitutes.

Proving Uniqueness. Because the game has strategic substitutes and is played over a compact action space, the set of Nash Equilibria is non-empty and forms a complete lattice. This guarantees the existence of a greatest equilibrium, $\boldsymbol{a}^{\max}$, and a least equilibrium, $\boldsymbol{a}^{\min}$, with respect to the standard vector ordering. We prove uniqueness by showing that these two equilibria must be the same.

Assume for contradiction that $\boldsymbol{a}^{\min} \neq \boldsymbol{a}^{\max}$. By the definition of least and greatest equilibria, we must have $\boldsymbol{a}^{\min} \leq \boldsymbol{a}^{\max}$ component-wise (i.e., $a_{kt}^{\min} \leq a_{kt}^{\max}$ for all k, t). Since they are distinct, there must exist at least one pair (n, t) for which the inequality is strict: $a_{nt}^{\min} < a_{nt}^{\max}$.

This ordering of actions implies an ordering of aggregate emissions, carbon concentrations, and temperatures for all periods $s \in \{1, \ldots, T\}$: $E_s^{\min} \leq E_s^{\max}$, $C_{s+1}^{\text{at,min}} \leq C_{s+1}^{\text{at,max}}$, and $\theta_{s+1}^{\text{at,min}} \leq \theta_{s+1}^{\text{at,max}}$.

Both $\boldsymbol{a}^{\min}$ and $\boldsymbol{a}^{\max}$ must satisfy the KKT conditions. For the pair (n, t), the condition $a_{nt}^{\min} < a_{nt}^{\max}$ implies (by complementary slackness) that $\overline{\lambda}_{nt}^{\min} = 0$ and $\underline{\lambda}_{nt}^{\max} = 0$. The first-order conditions are:

$$B_n'(a_{nt}^{\min}) + \underline{\lambda}_{nt}^{\min} = w_n \mathcal{M}_{nt}(\boldsymbol{a}^{\min}) \quad \text{and} \quad B_n'(a_{nt}^{\max}) - \overline{\lambda}_{nt}^{\max} = w_n \mathcal{M}_{nt}(\boldsymbol{a}^{\max}).$$

Subtracting the second equation from the first yields:

$$B_n'(a_{nt}^{\min}) - B_n'(a_{nt}^{\max}) + \underline{\lambda}_{nt}^{\min} + \overline{\lambda}_{nt}^{\max} = w_n \left(\mathcal{M}_{nt}(\boldsymbol{a}^{\min}) - \mathcal{M}_{nt}(\boldsymbol{a}^{\max}) \right).$$

By Assumption 1, B_n is strictly concave, so B_n' is strictly decreasing. Since $a_{nt}^{\min} < a_{nt}^{\max}$, the term $B_n'(a_{nt}^{\min}) - B_n'(a_{nt}^{\max})$ is strictly positive. As the Lagrange multipliers are non-negative, the left-hand side (LHS) is strictly positive. This implies the right-hand side (RHS) must also be strictly positive:

$$\mathcal{M}_{nt}(\boldsymbol{a}^{\min}) - \mathcal{M}_{nt}(\boldsymbol{a}^{\max}) > 0. \tag{B.5}$$

However, in the previous, we established that the condition on $g_s(\boldsymbol{E})$ implies that the game has strategic substitutes, meaning $\partial \mathcal{M}_{nt} / \partial a_{ks} \geq 0$ for all actions. This implies that the function $\mathcal{M}_{nt}(\boldsymbol{a})$ is non-decreasing in all of its arguments. Since $\boldsymbol{a}^{\min} \leq \boldsymbol{a}^{\max}$, it must be that:

$$\mathcal{M}_{nt}(\boldsymbol{a}^{\min}) \leq \mathcal{M}_{nt}(\boldsymbol{a}^{\max}).$$

This is equivalent to $\mathcal{M}_{nt}(\boldsymbol{a}^{\min}) - \mathcal{M}_{nt}(\boldsymbol{a}^{\max}) \leq 0$. This result contradicts (B.5), which established that the difference must be strictly positive.

The contradiction implies our initial assumption was false. Therefore, $\boldsymbol{a}^{\min} = \boldsymbol{a}^{\max}$, and the Nash Equilibrium is unique. □

Proof of (ii): Sufficiently Convex Damage Function. A unique Nash Equilibrium is guaranteed if the planning potential function, $\Phi(\boldsymbol{a})$, is strictly concave in the full strategy vector $\boldsymbol{a}$. The potential function is:

$$\Phi(\boldsymbol{a}; \boldsymbol{x}_1) = \left(\sum_{t=1}^{T} \sum_{n=1}^{N} \frac{B_n(a_{nt})}{w_n} \right) - \left(\sum_{t=1}^{T} D(\theta_{t+1}^{\text{at}}(\boldsymbol{a})) \right).$$

The first component, the sum of weighted benefits, is strictly concave in $\boldsymbol{a}$ as it is a sum of strictly concave functions of single variables (by Assumption 1).

Therefore, $\Phi(\boldsymbol{a})$ is strictly concave if the second component, the total discounted damage $-\sum_{t=1}^{T} D(\theta_{t+1}^{\text{at}}(\boldsymbol{a}))$, is concave in $\boldsymbol{a}$. Since the sum of concave functions is concave, it is sufficient to show that each term $-D(\theta_{r+1}^{\text{at}}(\boldsymbol{a}))$ is concave in $\boldsymbol{a}$ for any given $r \in \{1, \ldots, T\}$.

The proof for the static case (Proposition 8, part (iii)) established that the function $-D(\theta^{\text{at}})$ is concave with respect to the vector of single-period emissions $\boldsymbol{a}_r$ if the condition $D''(\theta^{\text{at}}) \geq D'(\theta^{\text{at}})/\beta_\theta$ holds. This proof relied on the fact that the temperature θ^{at} is a concave function of aggregate emissions $E_r = \sum_n a_{nr}$.

We can extend this logic directly. For any future period r, the temperature θ_{r+1}^{at} is a function of the history of aggregate emissions $(E_1, \ldots, E_r)$. The mapping from the full strategy vector $\boldsymbol{a}$ to this vector of aggregate emissions is linear. The reasoning from the static case proof shows that if $D''(\theta) \geq D'(\theta)/\beta_\theta$, the function $-D(\theta_{r+1}^{\text{at}})$ is a concave function of the aggregate emission history. Since the composition of a concave function with a linear map is concave, each term $-D(\theta_{r+1}^{\text{at}}(\boldsymbol{a}))$ is concave in the full strategy vector $\boldsymbol{a}$.

Thus, the total damage component is a sum of concave functions and is therefore concave. As the benefit component is strictly concave, the entire potential function $\Phi(\boldsymbol{a})$ is strictly concave. A strictly concave function on a compact convex set has a unique maximizer, which corresponds to the unique Pure-Strategy Nash Equilibrium. □

Appendix C. Proofs for Section 5

Appendix C.1. Proof of Proposition 11

Proof. The proof demonstrates that at the socially optimal profile $\boldsymbol{a}^{\mathrm{SO}}$, every player n has a unilateral incentive to increase their emissions in every time period.

A defecting player n solves the following optimization problem, taking the other players' cooperative actions $\boldsymbol{a}_{-n}^{\mathrm{SO}}$ as fixed:

$$\max_{\boldsymbol{a}_n \in \mathcal{A}_n^T} U_n(\boldsymbol{a}_n, \boldsymbol{a}_{-n}^{\mathrm{SO}}).$$

To determine the direction of deviation from the cooperative strategy $\boldsymbol{a}_n^{\mathrm{SO}}$, we analyze the gradient of player n's utility function with respect to their own strategy vector $\boldsymbol{a}_n$, evaluated at the social optimum $\boldsymbol{a}^{\mathrm{SO}}$. The components of this gradient, for each time period $t \in \mathcal{T}$, are:

$$\left.\frac{\partial U_n}{\partial a_{nt}}\right|_{\boldsymbol{a}=\boldsymbol{a}^{\mathrm{SO}}} = B_n'(a_{nt}^{\mathrm{SO}}) - w_n \mathcal{M}_{nt}(\boldsymbol{a}^{\mathrm{SO}}),$$

where $\mathcal{M}_{nt}(\boldsymbol{a}^{\mathrm{SO}}) = \sum_{s=t}^{T} \partial D(\theta_{s+1}^{\mathrm{at,SO}})/\partial a_{nt}$ is the marginal physical damage path evaluated at the social optimum.

From the first-order condition for an interior social optimum (Proposition 10), we have:

$$B_n'(a_{nt}^{\mathrm{SO}}) = \left(\sum_{k=1}^{N} w_k\right) \mathcal{M}_{nt}(\boldsymbol{a}^{\mathrm{SO}}).$$

Substituting this into the gradient expression yields:

$$\left.\frac{\partial U_n}{\partial a_{nt}}\right|_{\boldsymbol{a}=\boldsymbol{a}^{\mathrm{SO}}} = \left(\sum_{k=1}^{N} w_k\right) \mathcal{M}_{nt}(\boldsymbol{a}^{\mathrm{SO}}) - w_n \mathcal{M}_{nt}(\boldsymbol{a}^{\mathrm{SO}}) = \left(\sum_{k\neq n} w_k\right) \mathcal{M}_{nt}(\boldsymbol{a}^{\mathrm{SO}}).$$

Since $w_k > 0$ for all k by definition, and the marginal damage term $\mathcal{M}_{nt}(\boldsymbol{a}^{\mathrm{SO}})$ is strictly positive for an interior solution, every component of the gradient is strictly positive:

$$\left.\frac{\partial U_n}{\partial a_{nt}}\right|_{\boldsymbol{a}=\boldsymbol{a}^{\mathrm{SO}}} > 0 \quad \forall t \in \mathcal{T}.$$

This result shows that at the socially optimal point, player n can increase their total utility by marginally increasing their emissions in any time period. □

Appendix C.2. Proof of Proposition 13

Proof. From Proposition 11, we know that for any player n, it is optimal for them to unilaterally increase their emissions from the cooperative profile. To ensure that this incentive leads to a unique optimal defection $\boldsymbol{a}_n^D$ that is strictly greater than $\boldsymbol{a}_n^{\mathrm{SO}}$, the defector's objective function $U_n(\boldsymbol{a}_n, \boldsymbol{a}_{-n}^{\mathrm{SO}})$ must be strictly concave in the choice variable $\boldsymbol{a}_n$. The defector's utility is $U_n(\boldsymbol{a}_n, \boldsymbol{a}_{-n}^{\mathrm{SO}}) = \sum_{t=1}^{T} B_n(a_{nt}) - w_n \sum_{t=1}^{T} D(\theta_{t+1}^{\mathrm{at}}(\boldsymbol{a}_n, \boldsymbol{a}_{-n}^{\mathrm{SO}}))$. The benefit term is strictly concave in $\boldsymbol{a}_n$ by Assumption 1. The condition assumed in the proposition, $D''(\theta) \geq D'(\theta)/\beta_\theta$, is precisely the condition required to ensure that the damage component, $-w_n \sum_t D(\theta_{t+1}^{\mathrm{at}})$, is concave in $\boldsymbol{a}_n$ (as established in the proof of Proposition 9-(ii)).

Therefore, the defector's problem is the maximization of a strictly concave function over a convex set. For such a function, the fact that the gradient at point $\boldsymbol{a}_n^{\mathrm{SO}}$ is strictly positive in all components proves that the unique maximizer, $\boldsymbol{a}_n^D$, must be component-wise strictly greater than $\boldsymbol{a}_n^{\mathrm{SO}}$. □

References

[1] G. Hardin, The Tragedy of the Commons, Science 162 (3859) (1968) 1243–1248. doi:10.1126/science.162.3859.1243.

[2] W. D. Nordhaus, Managing the Global Commons: The Economics of Climate Change, MIT Press, Cambridge, MA, USA, 1994.

[3] H. Tulkens, Economics, Game Theory and International Environmental Agreements:The ca' Foscari Lectures, World Scientific Publishing Co. Pte. Ltd., 2019.

[4] C. Carraro, D. Siniscalco, Strategies for the international protection of the environment, Journal of Public Economics 52 (3) (1993) 309–328. doi:10.1016/0047-2727(93)90037-T.
[5] S. Barrett, Self-Enforcing International Environmental Agreements, Oxf Econ Pap 46 (Supplement_1) (1994) 878–894. doi:10.1093/oep/46.Supplement_1.878.
[6] C. Carraro, The structure of international environmental agreements, Fondazione Eni Enrico Mattei Work. Pap. (1998) 1–40.
[7] M. Finus, M. McGinty, Barrett's paradox of cooperation: A full analytical proof 30 years later, J. Environ. Econ. Manag. 128 (2024) 102960. doi:10.1016/j.jeem.2024.103045.
[8] C. D'Aspremont, A. Jacquemin, J. J. Gabszewicz, J. A. Weymark, On the Stability of Collusive Price Leadership, Can. J. Econ. Rev. Can. Econ. 16 (1) (1983) 17–25. arXiv:134972, doi:10.2307/134972.
[9] M. Finus, Game Theory and International Environmental Cooperation, Edward Elgar, Cheltenham, 2001.
[10] P. Chander, Game Theory and Climate Change, Columbia University Press, 2018. doi:10.7312/chan18464.
[11] W. D. Nordhaus, An optimal transition path for controlling greenhouse gases, Science 258 (5086) (1992) 1315–1319. doi:10.1126/science.258.5086.1315.
[12] W. D. Nordhaus, Z. Yang, A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies, Am. Econ. Rev. 86 (4) (1996) 741–765. arXiv:2118303.
[13] W. Nordhaus, Evolution of modeling of the economics of global warming: Changes in the DICE model, 1992–2017, Climatic Change 148 (4) (2018) 623–640. doi:10.1007/s10584-018-2218-y.
[14] IPCC, Chapter 3: Mitigation pathways compatible with long-term goals, in: Ipcc Ar6 Wgiii, 2022.
[15] IPCC, Annex III: Scenarios and modelling methods, Tech. rep., IPCC AR6 WGIII (2022).
[16] V. Bosetti, C. Carraro, E. De Cian, E. Massetti, M. Tavoni, Incentives and stability of international climate coalitions: An integrated assessment, Energy Policy 55 (2013) 44–56. doi:10.1016/j.enpol.2012.12.035.
[17] J. Emmerling, U. Kornek, V. Bosetti, K. Lessmann, Climate thresholds and heterogeneous regions: Implications for coalition formation, Rev Int Organ 16 (2) (2021) 293–316. doi:10.1007/s11558-019-09370-0.
[18] R. S. Pindyck, Climate Change Policy: What Do the Models Tell Us?, J. Econ. Lit. 51 (3) (2013) 860–872. doi:10.1257/jel.51.3.860.
[19] R. S. Pindyck, The Use and Misuse of Models for Climate Policy, Rev. Environ. Econ. Policy 11 (1) (2017) 100–114. doi:10.1093/reep/rew012.
[20] S. Dietz, N. Stern, Endogenous Growth, Convexity of Damage and Climate Risk: How Nordhaus' Framework Supports Deep Cuts in Carbon Emissions, The Economic Journal 125 (583) (2015) 574–620. doi:10.1111/ecoj.12188.
[21] M. Grubb, C. Wieners, P. Yang, Modeling myths: On DICE and dynamic realism in integrated assessment models of climate change mitigation, WIREs Clim. Change 12 (3) (2021) e698. doi:10.1002/wcc.698.
[22] F. Joos, G. Müller-Fürstenberger, G. Stephan, Correcting the carbon cycle representation: How important is it for the economics of climate change?, Environmental Modeling & Assessment 4 (2) (1999) 133–140. doi:10.1023/A:1019004015342.
[23] T. Faulwasser, R. Nydestedt, C. M. Kellett, S. R. Weller, Towards a FAIR-DICE IAM: Combining DICE and FAIR Models ∗∗TF acknowledges support from the Daimler Benz Foundation. TF, CMK, and SRW are supported by the Australian Research Council under ARC-DP180103026., IFAC-PapersOnLine 51 (5) (2018) 126–131. doi:10.1016/j.ifacol.2018.06.222.
[24] S. Dietz, F. van der Ploeg, A. Rezai, F. Venmans, Are Economists Getting Climate Dynamics Right and Does It Matter, J. Assoc. Environ. Resour. Econ. (2021). doi:10.1086/713977.
[25] M. L. Weitzman, On Modeling and Interpreting the Economics of Catastrophic Climate Change, The Review of Economics and Statistics 91 (1) (2009) 1–19. doi:10.1162/rest.91.1.1.
[26] M. L. Weitzman, Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change, Rev. Environ. Econ. Policy 5 (2) (2011) 275–292. doi:10.1093/reep/rer006.
[27] M. L. Weitzman, Fat Tails and the Social Cost of Carbon, Am. Econ. Rev. 104 (5) (2014) 544–546. doi:10.1257/aer.104.5.544.
[28] T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, H. J. Schellnhuber, Tipping elements in the Earth's climate system, Proc. Natl. Acad. Sci. 105 (6) (2008) 1786–1793. doi:10.1073/pnas.0705414105.
[29] T. M. Lenton, J. Rockström, O. Gaffney, S. Rahmstorf, K. Richardson, W. Steffen, H. J. Schellnhuber, Climate tipping points — too risky to bet against, Nature 575 (7784) (2019) 592–595. doi:10.1038/d41586-019-03595-0.
[30] S. Dietz, J. Rising, T. Stoerk, G. Wagner, Economic impacts of tipping points in the climate system, Proc. Natl. Acad. Sci. 118 (34) (2021) e2103081118. doi:10.1073/pnas.2103081118.
[31] F. van der Ploeg, A. J. de Zeeuw, International aspects of pollution control, Environ. Resour. Econ. 2 (2) (1992) 117–139. doi:10.1007/BF00338239.
[32] E. Calvo, S. J. Rubio, et al., Dynamic models of international environmental agreements: A differential game approach, Int. Rev. Environ. Resour. Econ. 6 (4) (2012) 289–339. doi:10.1561/101.00000053.
[33] A. de Zeeuw, Dynamic effects on the stability of international environmental agreements, J. Environ. Econ. Manag. 55 (2) (2008) 163–174. doi:10.1016/j.jeem.2007.06.003.
[34] B. Mroué, A. Couthures, S. Lasaulce, I.-C. Morărescu, On the impact of cross-country imitation on climate change: A game-theoretical analysis, IFAC Journal of Systems and Control 32 (2025) 100309. doi:10.1016/j.ifacsc.2025.100309.
[35] O. Geoffroy, D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon, S. Tytéca, Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments, J. Clim. 26 (6) (2013) 1841–1857. doi:10.1175/JCLI-D-12-00195.1.
[36] F. Joos, R. Roth, J. S. Fuglestvedt, G. P. Peters, I. G. Enting, W. von Bloh, V. Brovkin, E. J. Burke, M. Eby, N. R. Edwards, T. Friedrich, T. L. Frölicher, P. R. Halloran, P. B. Holden, C. Jones, T. Kleinen, F. T. Mackenzie, K. Matsumoto, M. Meinshausen, G.-K. Plattner, A. Reisinger, J. Segschneider, G. Shaffer, M. Steinacher, K. Strassmann, K. Tanaka, A. Timmermann, A. J. Weaver, Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: A multi-model analysis, Atmos. Chem. Phys. 13 (5) (2013) 2793–2825. doi:10.5194/acp-13-2793-2013.
[37] N. J. Leach, S. Jenkins, Z. Nicholls, C. J. Smith, J. Lynch, M. Cain, T. Walsh, B. Wu, J. Tsutsui, M. R. Allen, FaIRv2.0.0: A generalized

impulse response model for climate uncertainty and future scenario exploration, Geosci. Model Dev. 14 (5) (2021) 3007–3036. doi:10.5194/gmd-14-3007-2021.
[38] D. Monderer, L. S. Shapley, Potential Games, Games and Economic Behavior 14 (1) (1996) 124–143. doi:10.1006/game.1996.0044.
[39] Q. D. Lã, Y. H. Chew, B.-H. Soong, Potential Game Theory, Springer International Publishing, Cham, 2016. doi:10.1007/978-3-319-30869-2.
[40] D. H. Mguni, Y. Wu, Y. Du, Y. Yang, Z. Wang, M. Li, Y. Wen, J. Jennings, J. Wang, Learning in nonzero-sum stochastic games with potentials, in: M. Meila, T. Zhang (Eds.), Proc. 38th Int. Conf. Mach. Learn., Vol. 139 of Proceedings of Machine Learning Research, PMLR, 2021-07-18/2021-07-24, pp. 7688–7699.
[41] P. Chander, H. Tulkens, A core-theoretic solution for the design of cooperative agreements on transfrontier pollution, International tax and public finance 2 (2) (1995) 279–293.
[42] P. Chander, H. Tulkens, The core of an economy with multilateral environmental externalities, International Journal of Game Theory 26 (3) (1997) 379–401.
[43] J. R. Alaganthiran, M. I. Anaba, The effects of economic growth on carbon dioxide emissions in selected sub-saharan african (ssa) countries, Heliyon 8 (11) (2022).
[44] R. Winkler, On the relationship between adaptation and mitigation (2023).
[45] IPCC, Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change, Camb. Univ. Press (2013).
[46] IPCC, Climate change 2022: Mitigation of climate change. Contribution of working group III to the sixth assessment report of the intergovernmental panel on climate change, Camb. Univ. Press (2022).
[47] W. D. Nordhaus, The 'DICE' model: Background and structure of a dynamic integrated climate-economy model of the economics of global warming, Cowles Found. Discuss. Pap. (1009) (1992).
[48] R. S. J. Tol, The Economic Impacts of Climate Change, Rev. Environ. Econ. Policy 12 (1) (2018) 4–25. doi:10.1093/reep/rex027.
[49] J. Ayars, H. A. Kramer, G. M. Jones, The 2020 to 2021 California megafires and their impacts on wildlife habitat, Proc. Natl. Acad. Sci. 120 (48) (2023) e2312909120. doi:10.1073/pnas.2312909120.
[50] C.-C. Hong, A.-Y. Huang, H.-H. Hsu, W.-L. Tseng, M.-M. Lu, C.-C. Chang, Causes of 2022 Pakistan flooding and its linkage with China and Europe heatwaves, npj Clim Atmos Sci 6 (1) (2023) 163. doi:10.1038/s41612-023-00492-2.
[51] D. A. Driscoll, K. J. Macdonald, R. K. Gibson, T. S. Doherty, D. G. Nimmo, R. H. Nolan, E. G. Ritchie, G. J. Williamson, G. W. Heard, E. M. Tasker, R. Bilney, N. Porch, R. A. Collett, R. A. Crates, A. C. Hewitt, E. Pendall, M. M. Boer, J. Gates, R. L. Boulton, C. M. Mclean, H. Groffen, A. C. Maisey, C. T. Beranek, S. A. Ryan, A. Callen, A. J. Hamer, A. Stauber, G. J. Daly, J. Gould, K. L. Klop-Toker, M. J. Mahony, O. W. Kelly, S. L. Wallace, S. E. Stock, C. J. Weston, L. Volkova, D. Black, H. Gibb, J. J. Grubb, M. A. McGeoch, N. P. Murphy, J. S. Lee, C. R. Dickman, V. J. Neldner, M. R. Ngugi, V. Miritis, F. Köhler, M. Perri, A. J. Denham, B. D. E. Mackenzie, C. A. M. Reid, J. T. Rayment, A. Arriaga-Jiménez, M. W. Hewins, A. Hicks, B. A. Melbourne, K. F. Davies, M. E. Bitters, G. D. Linley, A. C. Greenville, J. K. Webb, B. Roberts, M. Letnic, O. F. Price, Z. C. Walker, B. R. Murray, E. M. Verhoeven, A. M. Thomsen, D. Keith, J. S. Lemmon, M. K. J. Ooi, V. L. Allen, O. T. Decker, P. T. Green, A. Moussalli, J. K. Foon, D. B. Bryant, K. L. Walker, M. J. Bruce, G. Madani, J. L. Tscharke, B. Wagner, C. R. Nitschke, C. R. Gosper, C. J. Yates, R. Dillon, S. Barrett, E. E. Spencer, G. M. Wardle, T. M. Newsome, S. A. Pulsford, A. Singh, A. Roff, K. J. Marsh, K. Mcdonald, L. G. Howell, M. R. Lane, R. H. Cristescu, R. R. Witt, E. J. Cook, F. Grant, B. S. Law, J. Seddon, K. K. Berris, R. M. Shofner, M. Barth, T. Welz, A. Foster, D. Hancock, M. Beitzel, L. X. L. Tan, N. A. Waddell, P. M. Fallow, L. Schweickle, T. D. Le Breton, C. Dunne, M. Green, A.-M. Gilpin, J. M. Cook, S. A. Power, K. Hogendoorn, R. Brawata, C. J. Jolly, M. Tozer, N. Reiter, R. D. Phillips, Biodiversity impacts of the 2019–2020 Australian megafires, Nature 635 (8040) (2024) 898–905. doi:10.1038/s41586-024-08174-6.
[52] R. Roson, Estimation of Climate Change Damage Functions for 140 Regions in the GTAP 9 Database, Journal of Global Economic Analysis 1 (2016) 78–115.
[53] S. Hsiang, R. Kopp, A. Jina, J. Rising, M. Delgado, S. Mohan, D. J. Rasmussen, R. Muir-Wood, P. Wilson, M. Oppenheimer, K. Larsen, T. Houser, Estimating economic damage from climate change in the United States, Science 356 (6345) (2017) 1362–1369. doi:10.1126/science.aal4369.
[54] D. Notz, The future of ice sheets and sea ice: Between reversible retreat and unstoppable loss, Proceedings of the National Academy of Sciences 106 (49) (2009) 20590–20595.
[55] E. A. Schuur, J. Bockheim, J. G. Canadell, E. Euskirchen, C. B. Field, S. V. Goryachkin, S. Hagemann, P. Kuhry, P. M. Lafleur, H. Lee, et al., Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle, BioScience 58 (8) (2008) 701–714.
[56] T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, H. J. Schellnhuber, Tipping elements in the earth's climate system, Proceedings of the national Academy of Sciences 105 (6) (2008) 1786–1793.
[57] O. L. De Silva, S. Lasaulce, I.-C. Morărescu, V. S. Varma, A game theory analysis of decentralized epidemic management with opinion dynamics, IEEE Transactions on Control of Network Systems 12 (1) (2025) 1137–1149.
[58] H. Ritchie, M. Roser, P. Rosado, CO2 and Greenhouse Gas Emissions, Our World in Data (May 2020).
[59] M. Roser, Future Population Growth, Our World in Data (May 2013).
[60] S. Waldhoff, D. Anthoff, S. Rose, R. S. J. Tol, The Marginal Damage Costs of Different Greenhouse Gases: An Application of FUND, Economics 8 (1) (Dec. 2014). doi:10.5018/economics-ejournal.ja.2014-31.
[61] A. Couthures, CO2-Game (Oct. 2025).
[62] H. E. Daly, For the common good: Redirecting the economy toward community, the environment, and a sustainable future, no. 73, Beacon Press, 1994.
[63] UNDP, Concept and measurement of human development, Human Development Report 1990 (1990).
[64] S. Managi, P. Kumar, Inclusive wealth report 2018, Taylor & Francis, 2018.
[65] K. Ura, S. Alkire, T. Zangmo, K. Wangdi, An extensive analysis of GNH index, 2012.
[66] J. Brizga, K. Feng, K. Hubacek, Drivers of CO2 emissions in the former Soviet Union: A country level IPAT analysis from 1990 to 2010, Energy 59 (2013) 743–753. doi:10.1016/j.energy.2013.07.045.

[67] U. Al-mulali, H. G. Fereidouni, J. Y. Lee, C. N. B. C. Sab, Examining the bi-directional long run relationship between renewable energy consumption and GDP growth, Renewable and Sustainable Energy Reviews 22 (2013) 209–222. doi:10.1016/j.rser.2013.02.005.

[68] D. P. van Vuuren, M. G. J. den Elzen, P. L. Lucas, B. Eickhout, B. J. Strengers, B. van Ruijven, S. Wonink, R. van Houdt, Stabilizing greenhouse gas concentrations at low levels: An assessment of reduction strategies and costs, Climatic Change 81 (2) (2007) 119–159. doi:10.1007/s10584-006-9172-9.

[69] R. N. Junichi Fujino, Multi-gas Mitigation Analysis on Stabilization Scenarios Using Aim Global Model, Energy J. Multi-Greenhouse Gas Mitigation and Climate Policy (Special Issue #3) (2006) 343–354. doi:10.5547/ISSN0195-6574-EJ-VolSI2006-NoSI3-17.

[70] K. Riahi, A. Grübler, N. Nakicenovic, Scenarios of long-term socio-economic and environmental development under climate stabilization, Technological Forecasting and Social Change 74 (7) (2007) 887–935. doi:10.1016/j.techfore.2006.05.026.

[71] W. Nordhaus, Estimates of the Social Cost of Carbon: Concepts and Results from the DICE-2013R Model and Alternative Approaches, J. Assoc. Environ. Resour. Econ. 1 (1/2) (2014) 273–312. doi:10.1086/676035.

[72] M. C. Hänsel, M. A. Drupp, D. J. A. Johansson, F. Nesje, C. Azar, M. C. Freeman, B. Groom, T. Sterner, Climate economics support for the UN climate targets, Nat. Clim. Chang. 10 (8) (2020) 781–789. doi:10.1038/s41558-020-0833-x.

[73] S. Keen, T. M. Lenton, T. J. Garrett, J. W. B. Rae, B. P. Hanley, M. Grasselli, Estimates of economic and environmental damages from tipping points cannot be reconciled with the scientific literature, Proc. Natl. Acad. Sci. 119 (21) (2022) e2117308119. doi:10.1073/pnas.2117308119.

[74] N. Bauer, L. Baumstark, M. Haller, J. Hilaire, M. Leimbach, G. Luderer, M. Lueken, R. Pietzcker, J. Strefler, S. Ludig, A. Koerner, A. Giannousakis, D. Klein, M. Pehl, C. Bertram, A. Schultes, REMIND:TheEquations.
URL https://www.pik-potsdam.de/en/institute/departments/transformation-pathways/models/remind

Appendix C

Résumé de la thèse en français

C.1 De l'impératif sociétal au défi scientifique

Le vingt-et-unième siècle traverse une phase décisive dans la lutte pour la stabilisation du climat terrestre. Alors que la température atmosphérique moyenne mondiale a déjà augmenté de plus de $1.3°C$ par rapport aux niveaux préindustriels, la question n'est plus de savoir *si* nous dépasserons la limite des $2°C$ fixée par l'Accord de Paris, mais *de combien* nous la dépasserons. Bien que l'humanité possède la compréhension scientifique et la capacité technologique nécessaires pour amorcer une transition vers un avenir bas-carbone, les progrès collectifs restent de manière alarmante insuffisants.

Cet écart grandissant entre l'ambition et les résultats met en évidence que la décarbonation ne peut se réduire à un défi technologique ou économique : il s'agit avant tout d'un problème *socio-technique*. La stagnation des politiques climatiques s'explique principalement par deux obstacles majeurs :

(i) **La compétition stratégique et économique :** La stabilité climatique est un bien public mondial. Le décalage entre les bénéfices collectifs et les coûts individuels encourage les comportements de passager clandestin (*free-riding*), conduisant à la classique « tragédie des biens communs ».

(ii) **Les dynamiques socio-comportementales :** Le succès des politiques climatiques dépend de l'acceptation sociale. Or, nos sociétés font face à une forte inertie comportementale et à une polarisation croissante, exacerbée par la structure des réseaux d'influence (chambres d'écho, désinformation).

Pour comprendre ces blocages, il est nécessaire de dépasser les modèles purement descriptifs pour aller vers une compréhension mécaniste. Cette thèse vise à combler les lacunes existantes entre les modèles climato-économiques (IAMs), la théorie des jeux et les modèles de dynamiques d'opinion. L'objectif est de construire une synthèse intégrant les choix stratégiques, les dynamiques non-linéaires d'influence sociale et les rétroactions d'un environnement en co-évolution.

C.2 Un parcours de recherche en trois étapes

L'analyse de ce système socio-environnemental complexe a été structurée en un programme de recherche en trois étapes, allant d'une approche macroscopique à une analyse fine des dynamiques non-linéaires sur les réseaux.

C.2.1 Interactions stratégiques et théorie des jeux

La première étape logique de nos travaux s'est attachée à formaliser la dimension stratégique de la décarbonation à l'échelle des nations. La question fondamentale qui a guidé cette phase était de déterminer s'il est possible de modéliser les prises de décision stratégiques des États en matière de politique climatique sous forme de jeu, afin de comprendre les conditions d'émergence d'accords climatiques stables et auto-réalisables (sans intervention extérieure). Pour y répondre, nous nous sommes appuyés sur le paradigme de la **théorie des jeux**, qui offre un cadre mathématique pour analyser les interactions entre plusieurs agents décisionnels via l'optimisation de fonctions d'utilité et la recherche d'**équilibres de Nash**.

État de l'art : une fracture disciplinaire. La littérature en modélisation des politiques climatiques s'est historiquement développée sur deux voies parallèles. D'une part, les **Modèles d'Évaluation Intégrée (IAMs)**, initiés par les travaux de Nordhaus (modèle DICE [17, 18]), offrent une vision systémique couplant économie et climat. Bien qu'indispensables pour l'analyse de scénarios, leur complexité en fait souvent des "boîtes noires", par ailleurs limitées par des modules climatiques initiaux simplifiés et des fonctions de dommages empiriquement fragiles et très sensibles au taux d'actualisation [9, 10, 34].

D'autre part, la **théorie des jeux environnementale** a privilégié une approche analytique pure [11–13]. Le modèle canonique représente l'utilité d'un agent comme un compromis entre les bénéfices de ses propres émissions et les dommages résultant de la somme des émissions mondiales. Si cette formulation abstraite permet d'obtenir des résultats mathématiques élégants, elle fait l'impasse sur la complexité géophysique (dynamiques non linéaires, inertie climatique). Cette abstraction a conduit à des conclusions théoriquement pessimistes, tel que le fameux "paradoxe de la coalition" [40], et a creusé un fossé entre la pureté analytique et la validation empirique.

Notre première contribution majeure consiste en une réponse directe à cette divergence, en réintégrant un modèle climatique dynamique au sein d'un cadre théorique des jeux mathématiquement traitable.

La construction d'un jeu socio-environnemental. Cette recherche s'est articulée autour de plusieurs étapes de complexification progressive, menées en collaboration avec le Dr. Bouchra Mroué :

– **Réintégration de la physique du climat :** Nous avons remplacé la fonction de dommages abstraite classique par des dommages dépendant de la **température atmosphérique globale (θ^{AT})**, issue d'un modèle climatique dynamique (SCM) couplé ([27–29]). Le résultat théorique central de cette formulation est la preuve que ce jeu constitue un **jeu de potentiel pondéré**. Cette propriété topologique est centrale : elle garantit l'existence d'équilibres de Nash purs et permet une recherche numérique efficace via la dynamique de meilleure réponse, évitant la présence d'une boîte noire.

– **Dérivation de solutions analytiques :** Dans un second temps, nous avons spécialisé ce modèle pour des fonctions de bénéfices et de dommages quadratiques. Cette étape a permis d'obtenir une **expression analytique** pour l'équilibre de Nash pur et unique. Cela a rendu les prédictions du modèle explicites, démontrant quantitativement comment la perception de la sévérité des dommages climatiques modifie directement les choix stratégiques des nations.

– **Mimétisme et réseaux mondiaux :** Les politiques climatiques n'évoluant pas en vase clos, nous avons enrichi le modèle en y intégrant un **terme d'imitation**, transformant l'interaction en un jeu sur réseau. Nous avons pu prouver formellement qu'une topologie de réseau plus connectée (un monde globalisé) conduit systématiquement à des températures d'équilibre plus

basses qu'un monde fragmenté en blocs concurrents, soulignant le rôle primordial de la diplomatie internationale et son devoir d'inclusion de tous les pays.

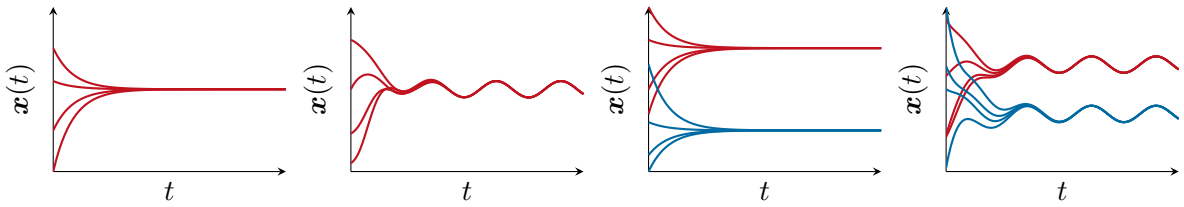
– **Planification et adaptation (L'horizon glissant) :** Enfin, nous avons étendu l'analyse myopique à la prise de décision sur un horizon temporel. Nous avons prouvé que la structure de jeu de potentiel est préservée dans un jeu de *planification multi-étapes*, et mis en évidence une propriété d'ordonnancement (supermodularité) des équilibres. Sur le plan politique, nos simulations révèlent que les fonctions de dommages quadratiques standard sont beaucoup trop optimistes pour induire la coopération nécessaire aux objectifs climatiques ; des dommages non-linéaires plus sévères sont requis. De plus, nous avons démontré que les politiques adaptatives basées sur une stratégie d'**horizon glissant** (*receding horizon*) génèrent des issues climatiques et des températures d'équilibre nettement inférieures aux engagements statiques à boucle ouverte.

Transition vers la dynamique des systèmes complexes. Si cette étude par la théorie des jeux nous a fourni un cadre pour identifier les points d'équilibre stratégiquement stables (les équilibres de Nash) et quantifier l'impact des leviers politiques, une limite fondamentale est apparue. L'analyse des états stationnaires ne permet pas de capturer la *dynamique temporelle du comportement collectif en lui-même*.

Elle reste opaque sur le comportement transitoire et ne peut expliquer l'émergence de phénomènes hors équilibre tels que les oscillations de l'opinion publique ou les basculements soudains de consensus politique. Cette prise de conscience a motivé un pivot conceptuel dans notre recherche : passer de l'analyse des **choix stratégiques statiques** à celle de l'**évolution dynamique des opinions couplées à un environnement en co-évolution**, marquant ainsi le début de la seconde phase de nos travaux.

C.2.2 Co-évolution Opinion-Environnement

Face aux limites inhérentes à l'analyse des équilibres de Nash, il est apparu important d'adopter une nouvelle perspective pour capturer l'évolution temporelle conjointe des opinions sociétales et de l'environnement. La question centrale de cette deuxième étape est alors devenue : *que se passe-t-il lorsque l'on couple explicitement les dynamiques sociales de formation d'opinion avec un état environnemental qui co-évolue ?*



(a) Consensus (b) Cons. oscillant (c) Polarisation (d) Pol. oscillante

Figure C.1: Comportements dynamiques observés du système couplé.

Pour y répondre, nous avons opéré une synthèse entre deux domaines : les **systèmes dynamiques non linéaires** (pour les concepts de stabilité, d'attracteurs et de bifurcations) et la **dynamique des opinions** (pour modéliser l'évolution des croyances sur des réseaux sociaux). Cette approche nous a permis d'observer quatre comportements dynamiques fondamentaux du système couplé : un consensus constant, un consensus oscillant, une polarisation constante, et une polarisation oscillante (voir Figure C.1).

État de l'art : L'essor des modèles socio-environnementaux couplés. Bien que l'idée d'une co-évolution entre société et environnement ne soit pas nouvelle, sa modélisation analytique a connu une accélération spectaculaire ces dernières années, suivant trois grands paradigmes :

– **Les jeux évolutifs et l'éco-évolution :** Initiée par des travaux en biologie mathématique [47, 48], cette branche a introduit la notion de "tragédie des communs oscillante", démontrant que la boucle de rétroaction peut générer des cycles auto-entretenus, ultérieurement explorés via la théorie des jeux à champ moyen ou la théorie du contrôle [49, 51].

– **L'approche épidémiologique :** Utilisant des modèles compartimentaux (ex. SIR), cette approche traite la propagation des comportements ou des rumeurs (comme le déni climatique) comme une contagion sociale [55, 56]. Son principal défaut réside dans l'utilisation de catégories comportementales discrètes (ex. croyant / sceptique), peinant à capturer le spectre continu et nuancé des opinions humaines.

– **Les modèles multi-agents (ABM) :** Issus de la physique sociale, ils simulent des règles individuelles avec une forte hétérogénéité d'agents [59, 60]. S'ils sont excellents sur le plan descriptif (par exemple pour modéliser l'impact du déni climatique [61, 62]), leur complexité entraîne souvent une perte de traçabilité analytique : ils montrent *que* le phénomène se produit, mais peinent à prouver mathématiquement *pourquoi*.

Le manque identifié : Au début de nos recherches, il manquait un cadre de modélisation capable de synthétiser les forces de ces approches. Il y avait un besoin pour un modèle fondé sur une **représentation continue de l'état d'opinion**, tout en étant suffisamment structuré pour permettre une **étude analytique rigoureuse** (via la théorie des bifurcations), sans sombrer dans l'intractabilité des simulations non traitables.

Nos contributions : Dévoiler les mécanismes des comportements dynamiques. Pour combler ce manque, nous avons développé et analysé un modèle couplé :

– **Exploration initiale en temps discret (Modèle CODA couplé) :** Nous avons d'abord couplé un modèle d'opinion continue à actions discrètes (CODA) avec une dynamique environnementale (ex. la qualité de l'air ou le climat). Notre principale découverte fut que le comportement du système dépend de la "force de couplage" β, c'est-à-dire la confiance accordée au signal environnemental. Si β est faible, le réseau s'auto-organise en clusters polarisés classiques et robustes. Mais au-delà d'un seuil critique ($\beta \geq 0.5$), ces équilibres sont détruits : le système entre dans un nouveau régime caractérisé par des oscillations persistantes (cycles limites) voire par un comportement **chaotique**. Cette étape a prouvé conceptuellement que la boucle de rétroaction altère fondamentalement la nature du système.

– **Analyse approfondie et réduction dimensionnelle :** En poussant l'analyse de ce système discret, nous avons pu établir l'**invariance en avant de la variété de synchronisation** (l'espace où tous les agents partagent la même opinion). En prouvant que les trajectoires démarrant sur cette variété y restent, nous avons pu réduire mathématiquement un problème complexe à N-dimensions à un système traitable à 2 dimensions (une opinion collective et un état environnemental), et dresser une cartographie systématique des espaces de paramètres.

– **Temps continu et Théorie des bifurcations :** Pour fournir une explication purement mécaniste, nous sommes passés à un modèle en temps continu. En utilisant le paramètre de confiance β comme paramètre de contrôle, nous avons pu prouver mathématiquement l'existence de deux transitions critiques :

(i) Une **bifurcation fourche** (*Pitchfork bifurcation*), qui marque le point où un consensus unifié perd sa stabilité et se scinde en deux états polarisés distincts, expliquant mécanistiquement l'émergence d'une bistabilité sociétale.

(ii) Une **bifurcation de Hopf**, où un équilibre stable cède la place à un cycle limite oscillant. C'est la preuve mathématique de l'origine des cycles d'attention publique, d'action et d'inaction climatique observés dans la société.

Bilan et transition vers l'étape finale. Cette deuxième phase a marqué un tournant critique dans notre étude. Elle a permis de dépasser les équilibres statiques et de prouver rigoureusement, via la caractérisation de bifurcations, non seulement *ce qui* se produit lors du couplage opinion-environnement, mais aussi *pourquoi* et *quand* les phénomènes d'oscillations et de polarisation apparaissent.

Cependant, une limitation demeurait : pour rendre cette analyse de bifurcation mathématiquement traitable, nous avons dû l'étudier sur la *variété de synchronisation* (où tous les agents finissent par s'accorder). Cette simplification laisse des questions cruciales en suspens : *que se passe-t-il dans un vaste réseau hétérogène où les agents ne se synchronisent pas complètement ? Comment les clusters polarisés se forment-ils, maintiennent-ils leur cohérence et interagissent-ils face à des signaux non linéaires ?* Pour y répondre, une étude approfondie des propriétés topologiques intrinsèques des modèles de dynamique d'opinion était nécessaire, ouvrant ainsi la voie à la troisième phase de notre recherche durant cette thèse.

C.3 Troisième étape : Les mécanismes de la polarisation

La découverte des dynamiques complexes lors de l'étape précédente a rendu impérative une caractérisation analytique rigoureuse. Si nous avions compris comment le couplage engendre des oscillations ou du chaos, il nous manquait une compréhension formelle des mécanismes régissant les transitions pures entre le consensus et la polarisation (l'axe horizontal de notre diagramme de phase). Cette troisième étape, qui constitue le cœur mathématique du manuscrit, répond à la question suivante : *Comment la topologie du réseau et la non-linéarité des communications dictent-elles la fragmentation ou l'unification d'une société ?*

État de l'art succinct.[7] Historiquement, les modèles fondateurs de la dynamique d'opinion (comme le modèle de DeGroot ou la dynamique Laplacienne [70, 71]) sont linéaires et prédisent inévitablement un consensus global sur un réseau connexe. Pour expliquer la polarisation observée dans la réalité, la littérature a dû introduire des mécanismes de rupture : l'entêtement des agents (modèles de Friedkin-Johnsen [83]), la confiance bornée (modèles de Krause [87]), ou les interactions antagonistes (graphes signés introduit par Altafini [90]).

Nos travaux s'inscrivent dans une quatrième voie, plus récente : **l'influence basée sur la perception et l'action**. Dans ce paradigme (qui inclut les approches par fonctions sigmoïdes [92, 96, 126] ou les modèles à actions discrètes (CODA) [67, 101]), l'influence ne s'exerce pas directement via l'opinion interne continue de l'agent, mais via un *signal* (une action publique, un vote) qui transforme cette opinion. C'est cette contrainte communicative qui devient la source mathématique du désaccord.

C.3.1 Le modèle COSA et la topologie de l'espace d'état (Chapitre 3)

Pour modéliser ce phénomène, nous avons développé le modèle (COSA) (*Continuous Opinion Signaled Action*), un cadre polyvalent qui généralise les modèles existants (notamment (CODA)) en autorisant une large classe de fonctions de communication non-linéaires. L'opinion interne

[7] Un état de l'art plus détaillé est fournis dans la partie en anglais de cette thèse.

continue d'un agent i, notée $x_i \in [-1, 1]$, évolue selon :

$$\dot{x}_i(t) = \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} a_{ij} s(x_j(t)) - x_i(t) \tag{C.1}$$

L'analyse dynamique et topologique de ce système nous a permis d'établir plusieurs propriétés fondamentales :

- **Coopérativité et absence de chaos :** En l'absence de rétroaction environnementale externe, le système (COSA) est intrinsèquement *coopératif* (son flux préserve l'ordre). Cette propriété garantit que presque toutes les trajectoires convergent vers un ensemble d'équilibres, excluant de facto les comportements oscillatoires ou chaotiques transitoires propres à l'Étape 2.
- **Structure globale de l'espace d'état :** Nous avons prouvé que l'hypercube global $\mathcal{X} = [-1, 1]^N$ est partitionné en sous-hypercubes positivement invariants. Fait remarquable, les frontières de ces hypercubes sont strictement définies par les points fixes instables de la fonction scalaire 1D $s(x)$. Ce résultat permet de localiser les trajectoires et de caractériser les bassins d'attraction dans un espace en grande dimension en se basant uniquement sur les propriétés de la fonction de signal.

C.3.2 Géométrie naturelle et seuil exact de synchronisation (Chapitre 4)

Sous quelles conditions le réseau parvient-il inévitablement à un consensus total, gommant ainsi toutes les conditions initiales ? L'obstacle mathématique à cette démonstration résidait dans la non-symétrie de l'opérateur d'interaction (la matrice d'adjacence normalisée par les degrés $\boldsymbol{D}^{-1}\boldsymbol{A}$).

Pour le surmonter, nous avons introduit une **géométrie naturelle** pour le système, basée sur un produit scalaire pondéré par les degrés : $\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\boldsymbol{D}} = \boldsymbol{u}^\top \boldsymbol{D}\boldsymbol{v}$. Sous cette métrique, l'opérateur redevient auto-adjoint, permettant une analyse de Lyapunov.

Ceci a conduit au résultat central de ce chapitre : la dérivation d'un seuil exact pour la synchronisation globale. Le système converge vers un consensus si et seulement si :

$$K\lambda_{N-1} < 1 \tag{C.2}$$

Cette équation offre une interprétation sociologique très claire. Elle met en balance :

- λ_{N-1} (la seconde plus grande valeur propre de la matrice d'itération), qui quantifie la **force de cohésion structurelle** dictée par la topologie du réseau.
- K (la constante de Lipschitz maximale du signal $s(x)$), qui représente le **potentiel de division** inhérent à la psychologie des agents et à la non-linéarité de la communication.

C.3.3 Théorie structurelle des clusters et robustesse (Chapitre 5)

Lorsque la condition de synchronisation échoue ($K\lambda_{N-1} \geq 1$), le système se fragmente et converge vers des Équilibres Non-Totalement Synchronisés (NFSE), c'est-à-dire vers la polarisation. Nous avons analysé la structure de ce désaccord selon deux paradigmes :

(i) **Clusters parfaits et réduction de modèle :** Nous avons démontré que la formation de clusters *parfaits* (où les agents d'un même groupe atteignent un accord numérique exact) n'est pas un comportement générique. Elle est strictement conditionnée par des symétries topologiques du graphe, spécifiquement l'existence d'une **partition régulière** (ou partition

équitable). Sous cette condition, le modèle peut être réduit exactement à un système quotient de dimension inférieure. Surtout, nous avons pu découpler mathématiquement la stabilité de ces clusters en deux conditions nécessaires et suffisantes indépendantes : la stabilité du *quotient* (qui régit la dynamique macroscopique inter-clusters) et la stabilité *transversale* (qui assure la cohésion microscopique intra-cluster).

(ii) **Clusters imparfaits et Stabilité Entrée-État (ISS) :** Les réseaux réels étant généralement dépourvus de symétries parfaites, nous avons développé un cadre novateur pour analyser la *synchronisation locale approximée*. En mobilisant les outils de la **Stabilité Entrée-État (ISS)** issus de la théorie du contrôle robuste, nous avons pu quantifier la résilience des clusters. Nous avons prouvé que le désaccord interne d'un cluster reste borné de manière robuste face aux influences du reste du réseau. Plus spécifiquement, l'analyse ISS révèle que ce n'est pas la force absolue de l'influence externe qui détruit un cluster, mais sa *variance*. Des signaux externes contradictoires vont fragmenter un groupe, tandis qu'une influence même extrême mais uniforme maintiendra sa cohésion interne, éclairant sous un angle mathématique la robustesse des *chambres d'écho*.

C.4 Conclusion générale

Cette thèse offre une boites à outils mathématique pour comprendre pourquoi et comment l'action collective, notamment face au changement climatique, peut échouer ou réussir. Le modèle (COSA), à travers l'analyse de Lyapunov en géométrie pondérée et l'approche ISS, démontre que la polarisation et la fragmentation sociale ne sont pas de simples anomalies comportementales, mais les résultats prévisibles et structurés de l'interaction entre la topologie d'un réseau de communication et les biais non-linéaires de ses acteurs. Ces résultats théoriques ouvrent de nouvelles perspectives, tant pour l'ingénierie du contrôle des systèmes multi-agents que pour le design des politiques publiques environnementales.

Bibliography

[1] Robert Vautard, Julien Cattiaux, Tamara Happé, Jitendra Singh, Rémy Bonnet, Christophe Cassou, Dim Coumou, Fabio D'Andrea, Davide Faranda, Erich Fischer, Aurélien Ribes, Sebastian Sippel, and Pascal Yiou. *"Heat Extremes in Western Europe Increasing Faster than Simulated Due to Atmospheric Circulation Trends"*. In: *Nat Commun* 14.1 (Oct. 2023), p. 6803 (cit. on p. 1).

[2] Chi-Cherng Hong, An-Yi Huang, Huang-Hsiung Hsu, Wan-Ling Tseng, Mong-Ming Lu, and Chih-Chun Chang. *"Causes of 2022 Pakistan Flooding and Its Linkage with China and Europe Heatwaves"*. In: *npj Clim Atmos Sci* 6.1 (Oct. 2023), p. 163 (cit. on p. 1).

[3] John T. Abatzoglou, Crystal A. Kolden, Alison C. Cullen, Mojtaba Sadegh, Emily L. Williams, Marco Turco, and Matthew W. Jones. *"Climate Change Has Increased the Odds of Extreme Regional Forest Fire Years Globally"*. In: *Nat Commun* 16.1 (July 2025), p. 6390 (cit. on p. 2).

[4] Jessalyn Ayars, H. Anu Kramer, and Gavin M. Jones. *"The 2020 to 2021 California Megafires and Their Impacts on Wildlife Habitat"*. In: *Proc. Natl. Acad. Sci.* 120.48 (Nov. 2023), e2312909120 (cit. on p. 2).

[5] Don A. Driscoll et al. *"Biodiversity Impacts of the 2019–2020 Australian Megafires"*. In: *Nature* 635.8040 (Nov. 2024), pp. 898–905 (cit. on p. 2).

[6] Katherine Calvin et al. *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]. IPCC, Geneva, Switzerland.* Tech. rep. Intergovernmental Panel on Climate Change (IPCC), July 2023 (cit. on p. 2).

[7] Garrett Hardin. *"The Tragedy of the Commons"*. In: *Science* 162.3859 (Dec. 1968), pp. 1243–1248 (cit. on p. 2).

[8] Aaron M. McCright and Riley E. Dunlap. *"The Politicization of Climate Change and Polarization in the American Public's Views of Global Warming, 2001–2010"*. In: *Sociol. Q.* 52.2 (May 2011), pp. 155–194 (cit. on p. 3).

[9] Robert S. Pindyck. *"Climate Change Policy: What Do the Models Tell Us?"* In: *J. Econ. Lit.* 51.3 (Sept. 2013), pp. 860–872 (cit. on pp. 3, 7, 216).

[10] Robert S. Pindyck. *"The Use and Misuse of Models for Climate Policy"*. In: *Rev. Environ. Econ. Policy* 11.1 (Jan. 2017), pp. 100–114 (cit. on pp. 3, 7, 216).

[11] Michael Finus. *"Game Theory and International Environmental Cooperation: Any Practical Application?"* Edward Elgar Publishing, 2002 (cit. on pp. 3, 7, 216).

[12] Parkash Chander. *"Game Theory and Climate Change"*. Columbia University Press, Dec. 2018 (cit. on pp. 3, 7, 216).

[13] Henry Tulkens. *"Economics, Game Theory and International Environmental Agreements:The ca' Foscari Lectures"*. World Scientific Publishing Co. Pte. Ltd., 2019 (cit. on pp. 3, 7, 8, 216).

[14] Paul-Marie Boulanger and Thierry Bréchet. *"Models for Policy-Making in Sustainable Development: The State of the Art and Perspectives for Research"*. In: *Ecological Economics* 55.3 (Nov. 2005), pp. 337–350 (cit. on p. 4).

[15] Chen Chris Gong, Behnaz Minooei Fard, Ibrahim Tahri, Chen Chris Gong, Behnaz Minooei Fard, and Ibrahim Tahri. *"Improving Accuracy, Complexity and Policy Relevance: A Literature Survey on Recent Advancements of Climate Mitigation Modeling"*. In: *AIMSES* 12.2 (2025), pp. 276–320 (cit. on p. 4).

[16] William D. Nordhaus. *"To Slow or Not to Slow: The Economics of The Greenhouse Effect"*. In: *The Economic Journal* 101.407 (July 1991), pp. 920–937 (cit. on p. 5).

[17] William D. Nordhaus. *"An Optimal Transition Path for Controlling Greenhouse Gases."* In: *Science* (1992) (cit. on pp. 5, 216).

[18] William D. Nordhaus and Zili Yang. *"A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies"*. In: *Am. Econ. Rev.* 86.4 (1996), pp. 741–765. JSTOR: 2118303 (cit. on pp. 5, 216).

[19] Erica L Plambeck and Chris Hope. *"PAGE95: An Updated Valuation of the Impacts of Global Warming"*. In: *Energy Policy* 24.9 (Sept. 1996), pp. 783–793 (cit. on p. 6).

[20] Stephanie Waldhoff, David Anthoff, Steven Rose, and Richard S. J. Tol. *"The Marginal Damage Costs of Different Greenhouse Gases: An Application of FUND"*. In: *Economics* 8.1 (Dec. 2014) (cit. on p. 6).

[21] Gunnar Luderer, Nico Bauer, Lavinia Baumstark, Christoph Bertram, Marian Leimbach, Robert Pietzcker, Jessica Strefler, Tino Aboumahboub, Gabriel Abrahão, Cornelia Auer, David Bantje, Felicitas Beier, Falk Benke, Stephen Bi, Jan Philipp Dietrich, Alois Dirnaichner, Tabea Dorndorf, Jakob Duerrwaechter, Sophie Fuchs, Pascal Führlich, Anastasis Giannousakis, Chen Chris Gong, Alex Hagen, Markus Haller, Robin Hasse, Léa Hayez, Jerome Hilaire, Verena Hofbauer, Johanna Hoppe, David Klein, Johannes Koch, Laurin Köhler-Schindler, Alexander Körner, Katarzyna Kowalczyk, Elmar Kriegler, Fabrice Lécuyer, Antoine Levesque, Alexander Lorenz, Sylvie Ludig, Michael Lüken, Aman Malik, Rahel Mandaroux, Susanne Manger, Anne Merfort, Leon Merfort, Simón Moreno-Leiva, Ioanna Mouratiadou, Jarusch Müßel, Adrian Odenweller, Michaja Pehl, Mika Pflüger, Franziska Piontek, Laura Popin, Sebastian Rauner, Oliver Richters, Renato Rodrigues, Niklas Roming, Ricarda Rosemann, Marianna Rottoli, Tonn Rüter, Robert Salzwedel, Pascal Sauer, Eva Schmidt, Christof Schötz, Felix Schreyer, Anselm Schultes, Joanna Sitarz, Björn Sörgel, Falko Ueckerdt, Philipp Verpoort, Pascal Weigmann, and Bennet Weiss. *REMIND - REgional Model of INvestments and Development*. Oct. 2025 (cit. on p. 6).

[22] Martin C. Hänsel, Moritz A. Drupp, Daniel J. A. Johansson, Frikk Nesje, Christian Azar, Mark C. Freeman, Ben Groom, and Thomas Sterner. *"Climate Economics Support for the UN Climate Targets"*. In: *Nat. Clim. Chang.* 10.8 (Aug. 2020), pp. 781–789 (cit. on p. 6).

[23] Rob Dellink, Jean Chateau, Elisa Lanzi, and Bertrand Magné. *"Long-Term Economic Growth Projections in the Shared Socioeconomic Pathways"*. In: *Global Environmental Change* 42 (Jan. 2017), pp. 200–214 (cit. on p. 6).

[24] F. Joos, G. Müller-Fürstenberger, and G. Stephan. *"Correcting the Carbon Cycle Representation: How Important Is It for the Economics of Climate Change?"* In: *Environmental Modeling & Assessment* 4.2 (May 1999), pp. 133–140 (cit. on p. 6).

[25] Mathijs J. H. M. Harmsen, Detlef P van Vuuren, Maarten van den Berg, Andries F Hof, Chris Hope, Volker Krey, Jean-Francois Lamarque, Adriana Marcucci, Drew T Shindell, and Michiel Schaeffer. *"How Well Do Integrated Assessment Models Represent Non-CO2 Radiative Forcing?"* In: *Climatic Change* 133.4 (Dec. 2015), pp. 565–582 (cit. on p. 6).

[26] Simon Dietz, Frederick van der Ploeg, Armon Rezai, and Frank Venmans. *"Are Economists Getting Climate Dynamics Right and Does It Matter"*. In: *J. Assoc. Environ. Resour. Econ.* (2021) (cit. on p. 6).

[27] F. Joos, R. Roth, J. S. Fuglestvedt, G. P. Peters, I. G. Enting, W. von Bloh, V. Brovkin, E. J. Burke, M. Eby, N. R. Edwards, T. Friedrich, T. L. Frölicher, P. R. Halloran, P. B. Holden, C. Jones, T. Kleinen, F. T. Mackenzie, K. Matsumoto, M. Meinshausen, G.-K. Plattner, A. Reisinger, J. Segschneider, G. Shaffer, M. Steinacher, K. Strassmann, K. Tanaka, A. Timmermann, and A. J. Weaver. *"Carbon Dioxide and Climate Impulse Response Functions for the Computation of Greenhouse Gas Metrics: A Multi-Model Analysis"*. In: *Atmos. Chem. Phys.* 13.5 (Mar. 2013), pp. 2793–2825 (cit. on pp. 6, 216).

[28] O. Geoffroy, D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon, and S. Tytéca. *"Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments"*. In: *J. Clim.* 26.6 (Mar. 2013), pp. 1841–1857 (cit. on pp. 6, 216).

[29] Nicholas Leach, Stuart Jenkins, Zebedee Nicholls, Chris Smith, John Lynch, Michelle Cain, Tristram Walsh, Bill Wu, Junichi Tsutsui, and Myles Allen. *"FaIRv2.0.0: A Generalized Impulse Response Model for Climate Uncertainty and Future Scenario Exploration"*. In: *Geoscientific Model Development* 14 (May 2021), pp. 3007–3036 (cit. on pp. 6, 216).

[30] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. *"Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization"*. In: *Geosci. Model Dev.* 9.5 (2016), pp. 1937–1958 (cit. on p. 6).

[31] Timm Faulwasser, Robin Nydestedt, Christopher M. Kellett, and Steven R. Weller. *"Towards a FAIR-DICE IAM: Combining DICE and FAIR Models ∗∗TF Acknowledges Support from the Daimler Benz Foundation. TF, CMK, and SRW Are Supported by the Australian Research Council under ARC-DP180103026."* In: *IFAC-PapersOnLine*. 1st IFAC Workshop on Integrated Assessment Modelling for Environmental Systems IAMES 2018 51.5 (Jan. 2018), pp. 126–131 (cit. on p. 6).

[32] Lint Barrage and William Nordhaus. *"Policies, Projections, and the Social Cost of Carbon: Results from the DICE-2023 Model"*. In: *Proc. Natl. Acad. Sci.* 121.13 (Mar. 2024), e2312030121 (cit. on p. 6).

[33] Martin Weitzman. *"Additive Damages, Fat-Tailed Climate Dynamics, and Uncertain Discounting"*. In: *Econ. - Open-Access Open-Assess. E-J.* 3 (Oct. 2009), pp. 1–29 (cit. on p. 7).

[34] Martin L. Weitzman. *"What Is the "Damages Function" for Global Warming — and What Difference Might It Make?"* In: *Clim. Change Econ.* 01.01 (May 2010), pp. 57–69 (cit. on pp. 7, 9, 216).

[35] Richard S. J. Tol. *"The Economic Impacts of Climate Change"*. In: *Rev. Environ. Econ. Policy* 12.1 (Jan. 2018), pp. 4–25 (cit. on p. 7).

[36] Richard S. J. Tol. *"Social Cost of Carbon Estimates Have Increased over Time"*. In: *Nat. Clim. Chang.* (May 2023), pp. 1–5 (cit. on p. 7).

[37] Clifford William Cobb. *"The Politics of Climate Denialism and the Secondary Denialism of Economics"*. In: *Am. J. Econ. Sociol.* 84.1 (2025), pp. 69–88 (cit. on p. 7).

[38] Valentina Bosetti, Carlo Carraro, Enrica De Cian, Emanuele Massetti, and Massimo Tavoni. *"Incentives and Stability of International Climate Coalitions: An Integrated Assessment"*. In: *Energy Policy* 55 (2013), pp. 44–56 (cit. on p. 7).

[39] Michael Finus, Francesco Furini, and Anna Viktoria Rohrer. *"The Efficacy of International Environmental Agreements When Adaptation Matters: Nash-Cournot vs Stackelberg Leadership"*. In: *J. Environ. Econ. Manag.* 109 (2021), p. 102461 (cit. on p. 7).

[40] Scott Barrett. *"Self-Enforcing International Environmental Agreements"*. In: *Oxf Econ Pap* 46.Supplement_1 (Oct. 1994), pp. 878–894 (cit. on pp. 7, 216).

[41] Bouchra Mroué. *"Decarbonization Strategies: A Game Theory Perspective"*. PhD thesis. Université de Lorraine, Apr. 2024 (cit. on p. 8).

[42] Bouchra Mroué, Anthony Couthures, Samson Lasaulce, and Irinel-Constantin Morarescu. *"Une Approche Par La Théorie Des Jeux Du Problème Du Choix Des Niveaux Nationaux d'émissions de CO2"*. In: *XXIXème Colloq. Francoph. Trait. Signal Images GRETSI 2023*. 2023 (cit. on pp. 8, 22).

[43] Anthony Couthures. *CO2-game* (cit. on pp. 8, 22).

[44] Dov Monderer and Lloyd S. Shapley. *"Potential Games"*. In: *Games and Economic Behavior* 14.1 (May 1996), pp. 124–143 (cit. on p. 8).

[45] Bouchra Mroué, Anthony Couthures, Samson Lasaulce, and Irinel Constantin Morărescu. *"A Quadratic Static Game Model for Assessing the Impact of Climate Change"*. In: *Complex Netw. Their Appl. XII*. Ed. by Hocine Cherifi, Luis M. Rocha, Chantal Cherifi, and Murat Donduran. Cham: Springer Nature Switzerland, 2024, pp. 383–394 (cit. on pp. 9, 22).

[46] Bouchra Mroué, Anthony Couthures, Samson Lasaulce, and Irinel-Constantin Morărescu. *"On the Impact of Cross-Country Imitation on Climate Change: A Game-Theoretical Analysis"*. In: *IFAC Journal of Systems and Control* 32 (June 2025), p. 100309 (cit. on pp. 9, 21).

[47] Joshua S. Weitz, Ceyhun Eksin, Keith Paarporn, Sam P. Brown, and William C. Ratcliff. *"An Oscillating Tragedy of the Commons in Replicator Dynamics with Game-Environment Feedback"*. In: *Proc. Natl. Acad. Sci.* 113.47 (Nov. 2016), E7518–E7525 (cit. on pp. 11, 218).

[48] Andrew R. Tilman, Joshua B. Plotkin, and Erol Akçay. *"Evolutionary Games with Environmental Feedbacks"*. In: *Nat Commun* 11.1 (Feb. 2020), p. 915 (cit. on pp. 12, 218).

[49] Leonardo Stella, Dario Bauso, and Patrizio Colaneri. *"Mean-Field Game for Collective Decision-Making in Honeybees via Switched Systems"*. In: *IEEE Trans. Automat. Contr.* 67.8 (Aug. 2022), pp. 3863–3878 (cit. on pp. 12, 218).

[50] Leonardo Stella and Dario Bauso. *"The Impact of Irrational Behaviors in the Optional Prisoner's Dilemma with Game-Environment Feedback"*. In: *Int. J. Robust Nonlinear Control* 33.9 (2023), pp. 5145–5158 (cit. on p. 12).

[51] Lulu Gong, Weijia Yao, Jian Gao, and Ming Cao. *"Limit Cycles Analysis and Control of Evolutionary Game Dynamics with Environmental Feedback"*. In: *Automatica* 145 (Nov. 2022), p. 110536 (cit. on pp. 12, 218).

[52] Hiromu Ito and Masato Yamamichi. *"A Complete Classification of Evolutionary Games with Environmental Feedback"*. In: *PNAS Nexus* 3.11 (Nov. 2024), pgae455 (cit. on p. 12).

[53] Yu-Hui Lin and Joshua S Weitz. *"Spatial Interactions and Oscillatory Tragedies of the Commons"*. In: *Phys. Rev. Lett.* 122.14 (2019), p. 148102 (cit. on p. 12).

[54] Katherine Betz, Feng Fu, and Naoki Masuda. *"Evolutionary Game Dynamics with Environmental Feedback in a Network with Two Communities"*. In: *Bull Math Biol* 86.7 (June 2024), p. 84 (cit. on p. 12).

[55] Athira Satheesh Kumar, Chris T. Bauch, and Madhur Anand. *"Climate-Denying Rumor Propagation in a Coupled Socio-Climate Model: Impact on Average Global Temperature"*. In: *PloS one* 20.1 (2025), e0317338 (cit. on pp. 12, 218).

[56] Ning Wang, Long Zhang, and Zhidong Teng. *"Dynamics in a Reaction-Diffusion Epidemic Model via Environmental Driven Infection in Heterogenous Space"*. In: *J. Biol. Dyn.* 16.1 (Dec. 2022), pp. 373–396 (cit. on pp. 12, 218).

[57] Ariful Kabir and K. M. *"Behavioral Vaccination Policies and Game-Environment Feedback in Epidemic Dynamics"*. In: *Sci Rep* 13.1 (Sept. 2023), p. 14520 (cit. on p. 12).

[58] Nataša Djurdjevac Conrad, Jonas Köppl, and Ana Djurdjevac. *"Feedback Loops in Opinion Dynamics of Agent-Based Models with Multiplicative Noise"*. In: *Entropy* 24.10 (Oct. 2022), p. 1352 (cit. on p. 13).

[59] Lorenzo Zino, Francesca Giardini, Daniele Vilone, and Ming Cao. *"On Modeling Collective Risk Perception via Opinion Dynamics"*. In: *European Journal of Control*. 2024 European Control Conference Special Issue 80 (Nov. 2024), p. 101036 (cit. on pp. 13, 218).

[60] Frances C. Moore, Katherine Lacasse, Katharine J. Mach, Yoon Ah Shin, Louis J. Gross, and Brian Beckage. *"Determinants of Emissions Pathways in the Coupled Climate–Social System"*. In: *Nature* 603.7899 (Mar. 2022), pp. 103–111 (cit. on pp. 13, 218).

[61] Tabea Hoffmann, Mengbin Ye, Lorenzo Zino, Ming Cao, Ward Rauws, and Jan Willem Bolderdijk. *"Overcoming Inaction: An Agent-Based Modelling Study of Social Interventions That Promote Systematic pro-Environmental Change"*. In: *Journal of Environmental Psychology* 94 (Mar. 2024), p. 102221 (cit. on pp. 13, 218).

[62] Kathinka Frieswijk, Lorenzo Zino, A. Stephen Morse, and Ming Cao. *"A Behavioural–Environmental Model to Study the Impact of Climate Change Denial on Environmental Degradation"*. In: *Physica D: Nonlinear Phenomena* 476 (June 2025), p. 134648 (cit. on pp. 13, 218).

[63] Carmela Bernardo, Lingfei Wang, Mathias Fridahl, and Claudio Altafini. *"Quantifying Leadership in Climate Negotiations: A Social Power Game"*. In: *PNAS Nexus* 2.11 (Nov. 2023), pgad365 (cit. on p. 13).

[64] Lorenzo Zino, Mengbin Ye, and Ming Cao. *"A Two-Layer Model for Coevolving Opinion Dynamics and Collective Decision-Making in Complex Social Systems"*. In: *Chaos Interdiscip. J. Nonlinear Sci.* 30.8 (Aug. 2020), p. 083107 (cit. on p. 13).

[65] K. Frieswijk, L. Zino, M. Cao, and A. S. Morse. *"Modeling the Co-evolution of Climate Impact and Population Behavior: A Mean-Field Analysis⋆"*. In: *IFAC-PapersOnLine*. 22nd IFAC World Congress 56.2 (Jan. 2023), pp. 7381–7386 (cit. on p. 13).

[66] Anthony Couthures, Thomas Mongaillard, Vineeth S. Varma, S. Lasaulce, and I.C. Morărescu. *"Analysis of a Continuous Opinion and Discrete Action Model Coupled with an External Dynamics"*. In: *2024 Eur. Control Conf. ECC*. June 2024, pp. 3392–3397 (cit. on pp. 14, 22, 145).

[67] N. R. Chowdhury, I.-C. Morărescu, S. Martin, and S. Srikant. *"Continuous Opinions and Discrete Actions in Social Networks: A Multi-Agent System Approach"*. In: *2016 IEEE 55th Conf. Decis. Control CDC*. Dec. 2016, pp. 1739–1744 (cit. on pp. 14, 19, 59, 61, 219).

[68] Anthony Couthures, Vineeth Satheeskumar Varma, Samson Lasaulce, and Irinel - Constantin Morărescu. *"Analysis of an Opinion Dynamics Model Coupled with an External Environmental Dynamics"*. In: *Chaos, Solitons & Fractals* 189 (Dec. 2024), p. 115719 (cit. on pp. 15, 21, 133).

[69] Anthony Couthures, Anastasia Bizyaeva, Vineeth S. Varma, Alessio Franci, and Irinel-Constantin Morarescu. *Bifurcation Analysis of an Opinion Dynamics Model Coupled with an Environmental Dynamics*. Apr. 2025. arXiv: 2504.03419 [eess] (cit. on pp. 15, 22, 23, 144).

[70] John R. P. French. *"A Formal Theory of Social Power."* In: *Psychological Review* 63.3 (1956), pp. 181–194 (cit. on pp. 17, 219).

[71] Morris H. Degroot. *"Reaching a Consensus"*. In: *J. Am. Stat. Assoc.* 69.345 (Mar. 1974), pp. 118–121 (cit. on pp. 17, 219).

[72] Reza Olfati-Saber, J. Alex Fax, and Richard M. Murray. *"Consensus and Cooperation in Networked Multi-Agent Systems"*. In: *Proc. IEEE* 95.1 (Jan. 2007), pp. 215–233 (cit. on p. 17).

[73] Roger L. Berger. *"A Necessary and Sufficient Condition for Reaching a Consensus Using DeGroot's Method"*. In: *J. Am. Stat. Assoc.* 76.374 (1981), pp. 415–418. JSTOR: 2287844 (cit. on p. 17).

[74] Mehran Mesbahi and Magnus Egerstedt. *"Graph Theoretic Methods in Multiagent Networks"*. STU - Student edition. Princeton University Press, 2010. JSTOR: j.ctt1287k9b (cit. on p. 17).

[75] Lin Xiao and Stephen Boyd. *"Fast Linear Iterations for Distributed Averaging"*. In: *Systems & Control Letters* 53.1 (Sept. 2004), pp. 65–78 (cit. on p. 17).

[76] A. Jadbabaie, Jie Lin, and A.S. Morse. *"Coordination of Groups of Mobile Autonomous Agents Using Nearest Neighbor Rules"*. In: *IEEE Trans. Autom. Control* 48.6 (June 2003), pp. 988–1001 (cit. on p. 17).

[77] Wei Ren and R.W. Beard. *"Consensus Seeking in Multiagent Systems under Dynamically Changing Interaction Topologies"*. In: *IEEE Trans. Automat. Contr.* 50.5 (May 2005), pp. 655–661 (cit. on p. 17).

[78] V.D. Blondel, J.M. Hendrickx, A. Olshevsky, and J.N. Tsitsiklis. *"Convergence in Multiagent Coordination, Consensus, and Flocking"*. In: *Proc. 44th IEEE Conf. Decis. Control*. Seville, Spain: IEEE, 2005, pp. 2996–3000 (cit. on pp. 17, 147).

[79] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. *"Statistical Physics of Social Dynamics"*. In: *Rev. Mod. Phys.* 81.2 (May 2009), pp. 591–646 (cit. on p. 17).

[80] Andreas Flache, Michael Mäs, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. *"Models of Social Influence: Towards the Next Frontiers"*. In: *JASSS* 20.4 (2017), p. 2 (cit. on p. 17).

[81] Anton Proskurnikov and Roberto Tempo. *"A Tutorial on Modeling and Analysis of Dynamic Social Networks. Part I"*. In: *Annual Reviews in Control* 43 (2017), pp. 65–79. arXiv: 1701.06307 [cs] (cit. on p. 17).

[82] Anton Proskurnikov and Roberto Tempo. *"A Tutorial on Modeling and Analysis of Dynamic Social Networks. Part II"*. In: *Annual Reviews in Control* 45 (2018), pp. 166–190. arXiv: 1801.06719 [cs] (cit. on p. 17).

[83] Noah E. Friedkin and Eugene C. Johnsen. *"Social Influence and Opinions"*. In: *J. Math. Sociol.* 15.3-4 (Jan. 1990), pp. 193–206 (cit. on pp. 18, 219).

[84] Noah E. Friedkin. *"The Problem of Social Control and Coordination of Complex Systems in Sociology: A Look at the Community Cleavage Problem"*. In: *IEEE Control Syst. Mag.* 35.3 (June 2015), pp. 40–51 (cit. on p. 18).

[85] Sergey E. Parsegov, Anton V. Proskurnikov, Roberto Tempo, and Noah E. Friedkin. *"Novel Multidimensional Models of Opinion Dynamics in Social Networks"*. In: *IEEE Trans. Autom. Control* 62.5 (May 2017), pp. 2270–2285 (cit. on p. 18).

[86] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. *"Mixing Beliefs among Interacting Agents"*. In: *Advs. Complex Syst.* 03.01n04 (Jan. 2000), pp. 87–98 (cit. on p. 18).

[87] R. Hegselmann and U. Krause. *"Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation"*. In: *J. Artif. Soc. Soc. Simul.* 5.3 (2002) (cit. on pp. 18, 219).

[88] Irinel-Constantin Morarescu and Antoine Girard. *"Opinion Dynamics With Decaying Confidence: Application to Community Detection in Graphs"*. In: *IEEE Trans. Autom. Control* 56.8 (Aug. 2011), pp. 1862–1873 (cit. on pp. 18, 135, 149).

[89] Claudio Altafini. *"Dynamics of Opinion Forming in Structurally Balanced Social Networks"*. In: *PLOS ONE* 7.6 (June 2012), e38135 (cit. on p. 18).

[90] Claudio Altafini. *"Consensus Problems on Networks With Antagonistic Interactions"*. In: *IEEE Trans. Autom. Control* 58.4 (Apr. 2013), pp. 935–946 (cit. on pp. 18, 219).

[91] Julien M. Hendrickx. *"A Lifting Approach to Models of Opinion Dynamics with Antagonisms"*. In: *53rd IEEE Conf. Decis. Control*. Los Angeles, CA, USA: IEEE, Dec. 2014, pp. 2118–2123 (cit. on p. 18).

[92] Rebecca Gray, Alessio Franci, Vaibhav Srivastava, and Naomi Ehrich Leonard. *"Multiagent Decision-Making Dynamics Inspired by Honeybees"*. In: *IEEE Trans. Control Netw. Syst.* 5.2 (June 2018), pp. 793–806 (cit. on pp. 19, 61, 219).

[93] Naomi Ehrich Leonard, Keena Lipsitz, Anastasia Bizyaeva, Alessio Franci, and Yphtach Lelkes. *"The Nonlinear Feedback Dynamics of Asymmetric Political Polarization"*. In: *Proc. Natl. Acad. Sci.* 118.50 (Dec. 2021), e2102149118 (cit. on p. 19).

[94] Alessio Franci, Martin Golubitsky, Anastasia Bizyaeva, and Naomi Ehrich Leonard. *A Model-Independent Theory of Consensus and Dissensus Decision Making*. Sept. 2020. arXiv: 1909.05765 [math] (cit. on p. 19).

[95] Alessio Franci, Martin Golubitsky, Ian Stewart, Anastasia Bizyaeva, and Naomi E. Leonard. *"Breaking Indecision in Multiagent, Multioption Dynamics"*. In: *SIAM J. Appl. Dyn. Syst.* 22.3 (Sept. 2023), pp. 1780–1817 (cit. on p. 19).

[96] Anastasia Bizyaeva, Alessio Franci, and Naomi Ehrich Leonard. *"Nonlinear Opinion Dynamics With Tunable Sensitivity"*. In: *IEEE Trans. Automat. Contr.* 68.3 (Mar. 2023), pp. 1415–1430 (cit. on pp. 19, 61, 147, 219).

[97] Anastasia Bizyaeva, Alessio Franci, and Naomi E. Leonard. *"Multi-Topic Belief Formation Through Bifurcations Over Signed Social Networks"*. In: *IEEE Trans. Autom. Control* (2025), pp. 1–16 (cit. on pp. 19, 147).

[98] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. *"Modeling Echo Chambers and Polarization Dynamics in Social Networks"*. In: *Phys. Rev. Lett.* 124.4 (Jan. 2020), p. 048301 (cit. on pp. 19, 147).

[99] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. *"Emergence of Polarized Ideological Opinions in Multidimensional Topic Spaces"*. In: *Phys. Rev. X* 11.1 (Jan. 2021), p. 011012 (cit. on p. 19).

[100] Łukasz G. Gajewski, Julian Sienkiewicz, and Janusz A. Hołyst. *"Transitions between Polarization and Radicalization in a Temporal Bilayer Echo-Chamber Model"*. In: *Phys. Rev. E* 105.2 (Feb. 2022), p. 024125 (cit. on pp. 19, 147).

[101] André C. R. Martins. *"Continuous Opinions and Discrete Actions in Opinion Dynamics Problems"*. In: *Int. J. Mod. Phys. C* 19.4 (2008), pp. 617–625 (cit. on pp. 19, 59, 61, 219).

[102] André C.R. Martins. *"Discrete Opinion Models as a Limit Case of the CODA Model"*. In: *Physica A: Statistical Mechanics and its Applications* 395 (Feb. 2014), pp. 352–357 (cit. on p. 19).

[103] Francesca Ceragioli and Paolo Frasca. *"Continuous-Time Consensus Dynamics with Quantized All-to-All Communication"*. In: *2015 Eur. Control Conf. ECC*. July 2015, pp. 1926–1931 (cit. on p. 19).

[104] F. Ceragioli and P. Frasca. *"Discontinuities, Generalized Solutions, and (Dis)Agreement in Opinion Dynamics"*. In: *Control Subject to Computational and Communication Constraints*. Ed. by Sophie Tarbouriech, Antoine Girard, and Laurentiu Hetel. Vol. 475. Cham: Springer International Publishing, 2018, pp. 287–309 (cit. on pp. 19, 147).

[105] Francesca Ceragioli and Paolo Frasca. *"Consensus and Disagreement: The Role of Quantized Behaviors in Opinion Dynamics"*. In: *SIAM J. Control Optim.* 56.2 (Jan. 2018), pp. 1058–1080 (cit. on pp. 19, 61, 75, 147).

[106] Raoul Prisant, Luca Cataldo, Francesca Ceragioli, and Paolo Frasca. *"Disagreement, Limit Cycles and Zeno Solutions in Continuous Opinion Dynamics with Binary Actions"*. In: *2024 Eur. Control Conf. ECC*. IEEE, 2024, pp. 2362–2367 (cit. on pp. 19, 68, 75).

[107] Francesca Ceragioli, Paolo Frasca, and Raoul Prisant. *"A Continuous Opinions Discrete Actions Model on Line Graphs"*. In: *J Optim Theory Appl* 206.2 (June 2025), p. 47 (cit. on p. 19).

[108] Anthony Couthures, Vineeth S. Varma, Samson Lasaulce, and Irinel-Constantin Morărescu. *"Global Synchronization of Multi-Agent Systems With Nonlinear Interactions"*. In: *IEEE Control Syst. Lett.* 9 (May 2025), pp. 354–359 (cit. on p. 21).

[109] Anthony Couthures, Prunelle Vogler, Olivier Beaude, Samson Lasaulce, and Irinel-Constantin Morarescu. *"An Optimization Setup of the Decarbonization Problem in the Transportation Sector"*. In: *Netw. Games Artif. Intell. Control Optim.* Ed. by Hélène Le Cadre, Yezekael Hayel, Bruno Tuffin, and Tijani Chahed. Cham: Springer Nature Switzerland, 2025, pp. 131–141 (cit. on pp. 22, 23).

[110] Anthony Couthures, Vineeth S. Varma, Samson Lasaulce, and Irinel-Constantin Morarescu. *Global Synchronization of Multi-Agent Systems with Nonlinear Interactions*. Mar. 2025. arXiv: 2503.10205 [eess] (cit. on p. 22).

[111] Anthony Couthures. *Private-Transport-Sector-Optimization-Model*. June 2024 (cit. on pp. 22, 23).

[112] Anthony Couthures. *COSA-Simulator*. Oct. 2025 (cit. on p. 23).

[113] Chris Godsil and Gordon Royle. *"Algebraic Graph Theory"*. Vol. 207. Graduate Texts in Mathematics. New York, NY: Springer New York, 2001 (cit. on pp. 26, 29, 38, 39).

[114] Francesco Bullo. *"Lectures on Network Systems"*. First edition. North Charleston, South Carolina: CreateSpace, 2018 (cit. on pp. 29, 31).

[115] Roger A. Horn and Charles R. Johnson. *"Matrix Analysis"*. Second edition, corrected reprint. New York, NY: Cambridge University Press, 2017 (cit. on pp. 32, 39, 40, 96).

[116] Delphine Bresch-Pietri and Laurent Praly. *"Fonctions de Lyapunov : Stabilité"*. Spartacus Supérieur. Spartacus-Idh, 2022, p. 396 (cit. on pp. 41–43, 52, 53).

[117] Hassan K Khalil. *"Nonlinear Systems"*. Upper Saddle River, N.J.: Prentice Hall, 2002 (cit. on pp. 42, 51, 53, 54, 124).

[118] Morris W. Hirsch. *"The Dynamical Systems Approach to Differential Equations"*. In: *Bull. New Ser. Am. Math. Soc.* 11.1 (July 1984), pp. 1–64 (cit. on p. 42).

[119] F. Blanchini. *"Set Invariance in Control"*. In: *Automatica* 35.11 (Nov. 1999), pp. 1747–1767 (cit. on p. 43).

[120] M.W. Hirsch and Hal Smith. *"Chapter 4 Monotone Dynamical Systems"*. In: *Handbook of Differential Equations: Ordinary Differential Equations*. Vol. 2. Elsevier, 2006, pp. 239–357 (cit. on pp. 46, 48, 50).

[121] Hal L. Smith. *"Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems"*. Mathematical Surveys and Monographs 41. American Mathematical Society, 2008 (cit. on pp. 47, 48, 50).

[122] T. H. Gronwall. *"Note on the Derivatives with Respect to a Parameter of the Solutions of a System of Differential Equations"*. In: *Ann. Math.* 20.4 (1919), pp. 292–296. JSTOR: 1967124 (cit. on p. 53).

[123] Andrzej Granas and James Dugundji. *"Fixed Point Theory"*. Springer Monographs in Mathematics. New York, NY: Springer New York, 2003 (cit. on p. 56).

[124] Paolo Frasca. *"Continuous-Time Quantized Consensus: Convergence of Krasovskii Solutions"*. In: *Syst. Control Lett.* 61.2 (2012), pp. 273–278 (cit. on p. 75).

[125] Hsiao-Dong Chiang and Luís F. C. Alberto. *"Stability Regions of Nonlinear Dynamical Systems: Theory, Estimation, and Applications"*. 1st ed. Cambridge University Press, June 2015 (cit. on p. 105).

[126] Angela Fontan and Claudio Altafini. *"Multiequilibria Analysis for a Class of Collective Decision-Making Networked Systems"*. In: *IEEE Trans. Control Netw. Syst.* 5.4 (Dec. 2018), pp. 1931–1940 (cit. on pp. 105, 106, 108, 146, 219).

[127] Jaroslaw Kwapisz. *"On the Spectral Radius of a Directed Graph"*. In: *J. Graph Theory* 23.4 (1996), pp. 405–411 (cit. on p. 126).

[128] F. Barahona. *"On the Computational Complexity of Ising Spin Glass Models"*. In: *J. Phys. A: Math. Gen.* 15.10 (Oct. 1982), p. 3241 (cit. on p. 127).

[129] Vaithilingam Jeyakumar. *"Nonsmooth Vector Functions and Continuous Optimization"*. Springer Optimization and Its Applications v. 10. New York: Springer, 2008 (cit. on pp. 129, 136).

[130] Wayne W. Zachary. *"An Information Flow Model for Conflict and Fission in Small Groups"*. In: *J. Anthropol. Res.* 33.4 (1977), pp. 452–473. JSTOR: 3629752 (cit. on p. 134).

[131] Ivan Penn, David Gelles Ivan Penn reported from Belém, São Paulo in Brazil, Pasadena, Calif., and David Gelles from Belém. *"Missing at U.N.'s Climate Meeting: American Executives"*. In: *The New York Times* (Nov. 2025) (cit. on p. 141).

[132] Omar Juarez-Alvarez and Alessio Franci. *"Collective Rhythm Design in Coupled Mixed-Feedback Systems Through Dominance and Bifurcations"*. In: *IEEE Trans. Control Netw. Syst.* 12.3 (Sept. 2025), pp. 1992–2003 (cit. on p. 144).

[133] Marcela Ordorica Arango, Anastasia Bizyaeva, Simon A. Levin, and Naomi Ehrich Leonard. *"Opinion-Driven Risk Perception and Reaction in SIS Epidemics"*. In: *2025 Am. Control Conf. ACC*. July 2025, pp. 4998–5003 (cit. on p. 144).

[134] Angela Fontan and Claudio Altafini. *"The Role of Frustration in Collective Decision-Making Dynamical Processes on Multiagent Signed Networks"*. In: *IEEE Trans. Autom. Control* 67.10 (Oct. 2022), pp. 5191–5206 (cit. on pp. 146, 150).

[135] F. Bullo. *"Contraction Theory for Dynamical Systems"*. 1.2. Kindle Direct Publishing, 2024 (cit. on p. 148).

[136] Bernard Brogliato, Bernhard Maschke, Rogelio Lozano, and Olav Egeland. *"Dissipative Systems Analysis and Control: Theory and Applications"*. Ed. by E. D. Sontag, M. Thoma, A. Isidori, and J. H. Van Schuppen. Communications and Control Engineering. London: Springer, 2007 (cit. on p. 148).

[137] Diego Romeres, Florian Dörfler, and Francesco Bullo. *"Novel Results on Slow Coherency in Consensus and Power Networks"*. In: *2013 Eur. Control Conf. ECC*. July 2013, pp. 742–747 (cit. on p. 150).

Résumé

Cette thèse modélise et analyse des systèmes socio-environnementaux complexes, une démarche motivée par l'urgence de l'action collective face au changement climatique. L'approche adoptée est multi-échelle (au sens spatial), allant du niveau stratégique macroscopique lors des négociations climatiques au niveau microscopique des dynamiques d'opinion individuelles. La thèse s'articule en trois axes principaux. Premièrement, la théorie des jeux est appliquée à la modélisation des négociations climatiques. Deuxièmement, l'étude du couplage entre la dynamique d'opinion et l'état de l'environnement révèle l'émergence de bifurcations et d'oscillations. Troisièmement, ce manuscrit se concentre sur l'analyse des dynamiques d'opinion sur des réseaux d'influence à travers le modèle que nous proposons : *Continuous Opinion Signaled Action* (COSA). Ce modèle capture des interactions non-linéaires où les agents communiquent via une fonction de signal. Nous démontrons les propriétés fondamentales du système en liant son analyse à celle d'un système unidimensionnel (*Chapitre 3*). Nous établissons les conditions de synchronisation complète et mettons en évidence un seuil spectral qui lie la topologie du réseau à la fonction de signal (*Chapitre 4*). Au-delà de ce seuil, nous caractérisons la formation de clusters d'opinion, que nous distinguons en deux catégories : parfaits et imparfaits. La synchronisation locale parfaite est conditionnée par des symétries structurelles du réseau. En l'absence de telles symétries, nous introduisons une nouvelle définition pour les clusters, dits imparfaits, et démontrons leur stabilité entrée-état face aux perturbations du reste du réseau (*Chapitre 5*).

Mots-clés : Dynamique d'opinion, Théorie des jeux, Théorie du contrôle, Changement climatique, Systèmes non-linéaires, Synchronisation, Formation de clusters (Clustering).

Abstract

This thesis models and analyzes complex socio-environmental systems, a research driven by the urgency of collective action in the face of climate change. The study is based on a multi-scale approach (in a spatial sense), progressing from a strategic macroscopic analysis of choices in climate negotiations to a microscopic analysis of individual opinion dynamics. The thesis is structured around three main axes. First, game theory is applied to model international climate negotiations. Second, the study of the coupling between opinion dynamics and the state of the environment reveals the emergence of bifurcations and oscillations. Third, this manuscript focuses on the analysis of opinion dynamics on influence networks through the model we introduce: the *Continuous Opinion Signaled Action* (COSA) model. This model captures nonlinear interactions where agents communicate via a signal function. We demonstrate the fundamental properties of the system by linking its analysis to that of a one-dimensional system (*Chapter 3*). We establish the conditions for full synchronization and identify a spectral threshold that connects the network topology to the signal function (*Chapter 4*). Beyond this threshold, we characterize the formation of opinion clusters, which we classify into two categories: perfect and imperfect. Perfect local synchronization is conditioned by structural symmetries of the network. In the absence of such symmetries, we introduce a new definition for imperfect clusters and demonstrate their input-to-state stability with respect to perturbations from the rest of the network (*Chapter 5*).

Keywords: Opinion dynamics, Game theory, Control theory, Climate change, Nonlinear systems, Synchronization, Clustering.