Εθνικο Μετσοβιο Πολυτεχνειο
Τμημα Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων

Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων
Εργαστηριο Λογικης και Επιστημης Υπολογιστων

Σύγκλιση και Σημεία Ισορροπίας σε Παίγνια Συνεξελικτικής Διαμόρφωσης Άποψης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Βασίλειου Α. Λίβανου

Επιβλέπων: Δημήτρης Φωτάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2017

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΛΟΓΙΚΗΣ ΚΑΙ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

Σύγκλιση και Σημεία Ισορροπίας σε Παίγνια Συνεξελικτικής Διαμόρφωσης Άποψης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Βασίλειου Α. Λίβανου

Επιβλέπων: Δημήτρης Φωτάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 21 Μαρτίου 2017.

.....	.....	.....
Δημήτρης Φωτάκης	Αριστείδης Παγουρτζής	Νικόλαος Παπασπύρου
Επίκουρος Καθηγητής Ε.Μ.Π.	Αναπληρωτής Καθηγητής Ε.Μ.Π .	Αναπληρωτής Καθηγητής Ε.Μ.Π .

Αθήνα, Μάρτιος 2017

.....

Βασίλειος Α. Λίβανος
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή αυτής της εργασίας, κ. Δημήτρη Φωτάκη, για την εμπιστοσύνη του, την αμέριστη βοήθεια και την συνεχή καθοδήγηση που μου προσέφερε. Ο ενθουσιασμός του για την μετάδοση της γνώσης και η ασταμάτητη ερευνητική του συνεισφορά υπήρξε για εμένα η μεγαλύτερη πηγή έμπνευσης και ο βασικότερος παράγοντας που με οδήγησε στο να ασχοληθώ ερευνητικά με το πεδίο της Θεωρητικής Πληροφορικής.

Επιπλέον, θα ήθελα να ευχαριστήσω όλα τα μέλη του εργαστηρίου Λογικής και Επιστήμης Υπολογιστών για την συνεχή υποστήριξη, φοβερή παρέα και τεράστια χαρά που μου έχουν προσφέρει όλα αυτά τα χρόνια. Θα ήθελα να ξεχωρίσω τον κ. Στάθη Ζάχο, για την ατελείωτη συμπαράστασή του. Η παρουσία και επίδρασή του πραγματικά κάνει το εργαστήριο την δεύτερη οικογένεια όλων και η φοιτητική μου πορεία δεν θα ήταν η ίδια χωρίς την επιρροή του. Επίσης, θα ήθελα να ευχαριστήσω τους κ. Άρη Παγουρτζή και κ. Νίκο Παπασπύρου για τις πολύτιμες γνώσεις και βοήθεια που προσφέρουν απλόχερα και οι δύο. Ακόμη, οφείλω πολλά στον Στρατή Σκουλάκη για τις άψογες συμβουλές και τις ατελείωτες ώρες συνεργασίας που έχουμε περάσει μαζί ψάχνωντας λύσεις σε ενδιαφέροντα προβλήματα.

Ιδιαίτερα θα ήθελα να ευχαριστήσω μέσα από την καρδιά μου τον Σωτήρη, τον Σεραφείμ, τον Παναγιώτη, τον Δημήτρη, τον Γιώργο, τον Κώστα, τον Νίκο και τον Στέργιο για τα καταπληκτικά φοιτητικά χρόνια και τις αξέχαστες εμπειρίες που περάσαμε μαζί. Η παρέα μας έζησε τις καλύτερες στιγμές και ειλικρινά δεν θα μπορούσα να φανταστώ αυτά τα χρόνια χωρίς την δική σας παρουσία.

Τέλος, θα ήθελα να εκφράσω ένα τεράστιο ευχαριστώ στην οικογενειά μου για την αμέριστη στήριξη και αγάπη που μου έχουν προσφέρει σε όλη την μέχρι τώρα πορεία μου. Ήταν και είναι πάντα δίπλα μου και χωρίς αυτούς δεν θα είχα καταφέρει τίποτα. Από την πρώτη στιγμή, η συνεισφορά τους στο να γίνω ο άνθρωπος που είμαι σήμερα είναι ανεκτίμητη.

Περίληψη

Στην σύγχρονη εποχή, οι ζωές μας επηρεάζονται πολύ από τα κοινωνικά δίκτυα στα οποία ανήκουμε. Παρόλαυτα, δεν έχουμε ακόμα κατανοήσει σε βάθος ούτε το πώς αυτά δουλεύουν, ούτε το πώς οι αλληλεπιδράσεις μεταξύ των ανθρώπων επηρεάζουν το κοινωνικό δίκτυο. Εμπνευσμένοι από αυτές τις παρατηρήσεις, προσπαθούμε να μοντελοποιήσουμε με μαθηματικό τρόπο τις αλληλεπιδράσεις και την ανταλλαγή απόψεων μεταξύ των πρακτόρων, εισάγωντας μοντέλα που προσομοιώνουν την διαμόρφωση των απόψεων στο κοινωνικό δίκτυο. Οι σημαντικότερες ερωτήσεις που προσπαθούμε να απαντήσουμε είναι εάν τα μοντέλα αυτά οδηγούν τους πράκτορες να συγκλίνουν σε συγκεκριμένες σταθερές απόψεις και, σε αυτήν την περίπτωση, πόσο γρήγορα το σύστημά μας φθάνει σε αυτή την σταθερή κατάσταση.

Σε αυτή την διπλωματική εργασία, μελετάμε σύνθετα μοντέλα που επιτρέπουν στις απόψεις των πρακτόρων και στο υποκείμενο κοινωνικό δίκτυο να συνεξελίσσονται. Αρχικά παρουσιάζουμε τα βασικά μοντέλα διαμόρφωσης άποψης, καθώς και τα σημαντικότερα αποτελέσματα για τις ιδιότητες σύγκλισής τους, και στην συνέχεια επικεντρώνουμε την ανάλυσή μας στο συνεξελικτικό μοντέλο Hegselmann-Krause (HK) και στις διάφορες παραλλαγές του. Συνεχίζουμε παρουσιάζοντας μία συλλογή από τα σημαντικότερα μαθηματικά εργαλεία και θεωρήματα που χρησιμοποιούνται για την ανάλυση πολλών μοντέλων διαμόρφωσης άποψης και, τελειώνοντας, καταδεικνύουμε την δύναμη της s-ενέργειας ενός συστήματος ως εργαλείο ανάλυσης, χρησιμοποιώντας το για την μελέτη των ιδιοτήτων σύγκλισης αρκετών παραλλαγών του HK μοντέλου.

Λέξεις-Κλειδιά: Αλγοριθμική Θεωρία Παιγνίων, Κοινωνικά Δίκτυα, Δυναμική Διαμόρφωση Άποψης, Συνεξελικτικά Μοντέλα

Abstract

In the modern world, our lives are heavily influenced by the social networks we belong to. However, we still do not have a deep understanding of the way they work, or how the interactions between agents influence the network. Inspired by this observations, we attempt to formalize the agents' interactions and exchange of opinion by introducing and analyzing several mathematical models that simulate the formation of opinions in our social network. The main questions we focus our attention on are whether these models lead the agents to converge to certain fixed opinions and, if so, how fast the system reaches this stable state.

In this diploma thesis, we study complex models that allow the agents' opinions and the underlying social network to coevolve. We begin by presenting the basic opinion formation models, along with the most important results about the convergence properties, before focusing on the Hegselmann-Krause (HK) coevolutionary model and its variants. We continue by introducing a collection of the most important mathematical tools and theorems that are used to analyze many opinion formation models and finally, we demonstrate the power of a system's s-energy as an analysis tool, as we use it to study the convergence properties of several variations of the HK model.

Keywords: Algorithmic Game Theory, Social Networks, Opinion Dynamics, Coevolutionary Models

Contents

1 Introduction **1**

1.1 Notation and Definitions 3

1.2 Organization of the Thesis 5

2 Opinion Formation Models **7**

2.1 Linear Models 7

2.1.1 The DeGroot Model 8

2.1.2 The Friedkin - Johnsen Model 12

2.2 Non-Linear (Coevolutionary) Models 14

2.2.1 The Hegselmann - Krause Model 15

2.2.2 The Deffuant - Weisbuch Model 18

3 Variations of the Hegselmann - Krause model **20**

3.1 The Network-HK Model 20

3.1.1 Definition 20

3.1.2 Results 21

3.2 The Random-HK Model 25

3.2.1 Definition 25

3.2.2 Results 26

3.3 The Inertial-HK Model 27

3.3.1 Definition 27

3.3.2 Results 28

3.4 The Asymmetric k-NN Model 35

3.4.1 Definition 35

3.4.2 Results 36

3.5 The Generalized Asymmetric Model . . . 37
3.5.1 Definition . . . 37
3.5.2 Results . . . 38

4 Model Analysis Toolbox **39**
4.1 Potential Functions . . . 39
4.1.1 Application to Opinion Dynamics . . . 41
4.2 Fixed-Point Theorems . . . 42
4.2.1 Brouwer's Fixed-Point Theorem . . . 43
4.2.2 Kakutani's Fixed-Point Theorem . . . 45
4.3 Concave Games . . . 46
4.3.1 Rosen's Theorem . . . 47
4.4 Gradient Descent Methods . . . 49
4.4.1 Gradient Descent . . . 50
4.4.2 Mirror Descent . . . 54
4.5 Energy as a Generating Function . . . 55
4.5.1 Definition of the Total s-Energy . . . 55
4.5.2 Bidirectional Systems . . . 56
4.5.3 Bounds on the Total s-Energy . . . 57
4.5.4 Kinetic s-Energy . . . 58

5 Convergence of Variations of the Hegselmann - Krause Model **60**
5.1 Energy Approach to the Network-HK Model . . . 60
5.2 Analysis of the Inertial-HK Model . . . 64
5.3 Future Work . . . 68

List of Figures

2.1 An example of a convergent DeGroot model 9

2.2 An example of a non-convergent DeGroot model 11

4.1 An example of Kakutani's fixed-point theorem 46

Chapter 1

Introduction

In this diploma thesis, we are dealing with opinion formation models in social networks. The complexity and significance of real-world social networks in our everyday lives makes the need to understand them more important than ever. This need is amplified by the rapid growth of the Internet which provides the medium for larger, more complex social networks to exist, that are quite different from traditional social networks and lead to an even more dynamic opinion exchange between the people in the network. The high level of correlation between the interactions in real social networks lead us to develop mathematical models to aid in our analysis of their behavior. Hence, we are able to study these models in a mathematical framework, and provide concrete proofs about their interesting properties.

Opinion formation models are of significant importance, as their applications can be observed all around us, at any interaction we take part in. The underlying principles that govern the way these models work attempt to explain our behavior in real social networks, and any information we can discover about their properties will have an immediate effect on the way we choose our opinions about a certain topic. Observations about how people form their opinions can impact fields ranging from psychology and sociology to political science and social choice. Almost all people hold an opinion about varying factors, like economic welfare, religion, ecucation and culture, and these opinions can only be transmitted from one person to another. Since the disagreement of a person with the other people in his social group incurs upon them some kind of cost, from psychological effects to even social ostracisation, studying such interactions has significant merit, if we wish to understand why people all around the world behave as they do.

Due to their ubiquity and importance, opinion dynamics have been studied extensively [1]. While the study of opinion formation models is a field that exists for quite

some time, its recent revival and the use of modern techniques has allowed for the discovery of several key models and properties, while providing us with the most significant results as of yet. The study of such concepts began in 1956, with the ideas presented by French [2], but the first important results about the convergence properties of early models were provided by the pioneering work of DeGroot in 1974 [3]. The next decades saw the research community attempting to generalize DeGroot's simplistic model, which led to the development of many variations, the most important of them being the Friedkin-Johnsen model [4]. However, we were not yet satisfied by the complexity of interactions that our models could simulate, and that line of thinking led to the development of *coevolutionary* models, with the most important one being the Hegselmann-Krause model [5]. More recent attempts to model real social networks include randomness in the way that people in the network update their opinions [6, 7].

While we wish to understand as much as we can about every model that describes certain social networks, we usually focus on two main questions. Specifically, we attempt to discern whether our models lead the people in our network to converge to a specific set of fixed opinions and, if so, how fast they approach these opinions. Our approach is entirely algorithmic in its nature and we view such models from a game-theoretic perspective, where we study the macroscopic effect of the selfish actions of people in our network. Motivated by observations similar to ours, the convergence properties of opinion formation models have been extensively studied in the discrete setting, where people choose from a finite set of discrete opinions [8–10]. This line of thiniking is closer in spirit to the fields of social choice and voting theory.

Other approaches focus on the importance of going beyond understanding when agents converge to fixed opinions, and studying the *social cost* of the outcomes that emerge [11–13]. These studies are generally inspired by the observation that convergence rarely emerges in real social networks and, even if it does, different equilibria have different levels of desirability among the people. Recent work attempts to blend the field of opinion dynamics with other, seemingly unrelated, fields such as chemotaxis, synchronization and bird flocking, with the introduction of systems that generalize all the aforementioned fields, called *influence systems* [14–17]. Such systems are observed in nature, and typically consist of a general rule that indicates how each *agent* in the system updates his state by processing the information it receives from its neighbors.

Our focus in this thesis lies in coevolutionary models, in which the peoples' opinions and the underlying social network continuously affect each other, thus they coevolve. Other approaches study coevolutionary generalizations of the DeGroot model [18], or models in which agents are skeptical twoards opinions far away from theirs [19]. We study the Hegselmann-Krause model in the continuous opinion space, introduce several of its variations and attempt to analyze their convergence properties. However, we feel

that our approach is not complete, unless we also provide a collection of useful mathematical tools utilized to derive significant results about these models. The complexity of coevolutionary models has shown the need to associate techniques used in different areas of algorithmic game theory, or even derive new ones. Theories such as potential functions, fixed-point theorems and concave games have shed light on the existence and uniqueness of stable states in our system, while tools like gradient descent and the, newly invented, s-energy of a system are very helpful in our attempts to discern which conditions lead opinion formation models to converge to these stable states.

We begin by providing some necessary definitions and notation, followed by an overview of the information presented in each chapter.

1.1 Notation and Definitions

In this section, we are going to present the basic notation and definitions required to follow the ideas presented in the rest of this thesis. In every opinion formation model there exists a group of n agents, usually numbered 1 to n, that express their opinions and beliefs about certain topics. We represent these expressed opinions by a vector $\boldsymbol{x}$, where x_i denotes the expressed opinion of agent i. Usually we study the models in 1 dimension, hence $x_i \in \mathbb{R}$ or even $x_i \in [0, 1]$ in some cases. In the most general setting however, each agent expresses an opinion about multiple topics, simultaneously, and each x_i is a tuple consisting of these opinions. Therefore, in this case, $x_i \in \mathbb{R}^d$, where the model's dimension d is the number of simultaneous opinions that each agent expresses. For example, these opinions can represent i's position on the political spectrum, i's sympathy towards a specific sports team, or even a probability that i holds a particular belief.

Furthermore, in every model, each agent forms at each step a *neighborhood* that determines which other agents influence her opinion. We denote i's neighborhood by $\mathcal{N}_i$, and note that it can vary with time, or depend on other model-specific parameters. In cases where $|\mathcal{N}_i| = n$ for every agent i, we have an instance of the model where each agent influences all others, and information and opinions spread quickly through our network. When $\mathcal{N}_i = \{i\}$, we understand that agent i is influenced only by his own opinion, therefore we expect her to remain at a fixed position.

We also define the concept of the *pure strategy Nash equilibrium*, introduced in the field of game theory by Nash [20], which, when applied to opinion dynamics, clearly describes a state where no agent has an interest in changing her opinion. The Nash equilibrium has emerged as the standard solution concept in non-cooperative games, and opinion formation models are no exception. Indeed, our analysis usually consists

of first attempting to prove the existence of such an equilibrium for a given model, and then trying to show that the dynamics of our model converge to an equilibrium.

Next, we define a special kind of equilibrium called *consensus*.

Definition 1.1 (Consensus)**.** Consider an opinion formation model with n agents where all agents share exactly the same opinion. Therefore, for all agents i

$$x_i = x^* \tag{1.1}$$

We note that no agent has an incentive to deviate from x^*, therefore this state is a Nash equilibrium. We call this equilibrium where all agents express the same opinion a *consensus*.

As stated before, we study models from the aspect of their convergence to an equilibrium. However, another inetresting question to consider here is, given that the system converges to a specific equilibrium, how quickly do agents approach their opinions at that equilibrium? This question demonstrates the need for formalization of the concept of the number of time steps it takes for agents' opinions to converge. In our models, we call this number the *rate of convergence* and, since our approach is algorithmic, we are interested only in its asymptotic behavior.

Continuing with our definitions, the *update rule* of a model is the rule that governs the way each agent updates her expressed opinion at each time step. Usually, we denote such a rule as

$$\boldsymbol{x}(t+1) = \boldsymbol{A}(\boldsymbol{x}(t))\boldsymbol{x}(t) \tag{1.2}$$

in the general setting. However, there are many variations of this rule, as we will present in the following chapters, and each one defines a different opinion formation model. $\boldsymbol{A}(\boldsymbol{x}(t))$ is a $n \times n$ matrix which encodes the influence that any agent has towards any other. While $\boldsymbol{A}(\boldsymbol{x}(t))$ can generally be any matrix, the logical and useful models impose some constraints on agent interactions, therefore on $\boldsymbol{A}(\boldsymbol{x}(t))$ as well.

In many models we are given a weighted graph $G = (V, E)$, which represents the underlying social network. In G, every node represents an agent, and the existence and weight of an edge between two nodes represents the strength of influence between the two corresponding agents.

Next up, we characterize the agents according to their willingness to change opinions, or *stubbornness*. We distinguish agents having $w_{ii} > 0$, who we call *stubborn agents*, and agents having $w_{ii} = 0$, who we call *non-stubborn agents*. We can distinguish even further

among stubborn agents, between those having $w_{ii} = 1$ and $w_{ij} = 0$, for any other agent j, who we call *fully-stubborn agents*, and those having $w_{ii} < 1$, who we call *partially-stubborn agents*.

It is therefore evident that these definitions can model a wide variety of phenomena in which nodes in a system hold a numerical value, influence each other and update their values. In these phenomena the agents need not necessarily be people, but could also be computers in a network, as is sometimes the case.

1.2 Organization of the Thesis

In *Chapter 2*, we introduce the basic opinion formation models. We distinguish between linear and non-linear (coevolutionary) models, and begin by presenting the DeGroot model, perhaps the most important among the linear models. We provide its definition and define clearly which properties lead to convergence when they exist. We continue by introducing the Friedkin-Johnsen (FJ) model, a variation of DeGroot's, and state its known convergence results. Next, we switch our attention to coevolutionary models, and introduce the most significant among them, the Hegselmann-Krause model. We present its characteristics and convergence properties, before moving on to the Deffuant-Weisbuch model, which differs from the all models mentioned before, as it introduces randomness in the agents' update rule.

The purpose of *Chapter 3* is to introduce several variants of the Hegselmann-Krause model. In the Network-HK model, we extend the original HK model with the addition of an underlying graph which imposes constraints on each agent's neighborhood, thus limiting the spread of information through the social network. Next, we present the Random-HK model, which is an attempt to introduce randomness to the original HK model. We continue with the Inertial-HK model, which is a heterogeneous variant of the original HK, and use it to settle the issue of convergence for the HK model with fully-stubborn agents. In the Asymmetric k-Nearest Neighbor (k-NN) model, we present, for the first time, a variant of the HK model that does not always converge and provide an indicative counterexample. Finally, we introduce the Generalized Asymmetric model, an attempt to provide a much more general opinion formation model, with unknown convergence properties as of yet. However, we show that it always admits to a pure strategy Nash equilibrium, under certain assumptions.

In *Chapter 4*, we present a collection of mathematical tools and theorems utilized in the analysis of several different opinion formation models. We distignuish between tools used to prove the existence, and sometimes uniqueness, of equilibria and tools used to show whether a system converges or not. The former include potential functions, fixed-point theorems and concave functions, from which we derive an important result which

characterizes the models that admit to equilibria. The latter include gradient descent methods and s-energy methods, with the last one being a newly introduced concept in our field which may yet provide interesting results.

Chapter 5 contains our research work. Initially, we provide a proof of convergence for the Network-HK model using the s-energy approach. Then, we attempt to provide some intuition into the proof of convergence of the Inertial-HK model, as we believe the thought process hidden behind the proof is of substantial value.

Chapter 2

Opinion Formation Models

The purpose of this chapter is to introduce the reader to the most commonly studied and analyzed opinion formation models in the field of opinion dynamics. In addition to the basic definitions, we present a collection of the most important algorithms and results. We present the models starting from simple, linear models and progressing to more complex ones, that capture a wider variety of real-world scenarios.

2.1 Linear Models

In every linear model there exists a group of n agents, numbered 1 to n, that express their opinions and beliefs about a certain topic. We assume that each agent i holds an opinion that is equal to a real number x_i. Furthermore, there exists a weighted graph $G = (V, E)$, that represents the underlying social network. Each agent is represented by a node in G, and agent i is influenced by agent j's opinion if and only if there exists an edge $e = (i, j)$ in G, from i to j. Moreover, the weight of the edge e, denoted by $w_{ij} \geq 0$, represents the strength of j's influence over i.

Before we proceed, we will provide some necessary notation. We will denote the vector of the agents' opinions by $\boldsymbol{x}(t)$ at time t and the matrix of the agents' weights by $\boldsymbol{A}$, meaning that the (i, j) element of matrix $\boldsymbol{A}$ is w_{ij}. In linear models, the influence of one agent over another does not change with time. Therefore, the underlying graph G is a steady graph, and $\boldsymbol{A}$ is a constant matrix. If there does not exist an edge from agent i to agent j in G, we assign $w_{ij} = 0$ to capture the fact that i's opinion does not depend on j's opinion. We also define the set $\mathcal{N}_i = \{j : (i, j) \in E(G)\}$ and call it the *neighborhood* of agent i. When our model allows for self loops in G, we assume $w_{ii} \neq 0$ and it is understood that $i \in \mathcal{N}_i$ for every agent i.

2.1.1 The DeGroot Model

The first model that we introduce is a very simple yet incredibly expressive model first introduced by DeGroot [3] in 1974. In the *DeGroot model*, at each time step t, every agent i updates her opinion to the weighted average of her current opinion and the current opinions of her neighbors. Without loss of generality, we assume that $\sum_{j \in \mathcal{N}_i} w_{ij} = 1$, since we can always normalize the weights to 1. This means that $\boldsymbol{A}$ is a stochastic matrix, since each row of $\boldsymbol{A}$ consists of nonnegative real numbers, and sums up to 1. It is easy to see that

$$x_i(t+1) = w_{ii}x_i(t) + \sum_{\substack{j \in \mathcal{N}_i \\ j \neq i}} w_{ij}x_j(t) \tag{2.1}$$

or, in matrix form

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) \tag{2.2}$$

We can distinguish between two cases of the DeGroot model that give different results, the *undirected* and the *directed* DeGroot model. In the undirected case, we assume that for every pair of agents i and j, $w_{ij} = w_{ji}$, which means that the influence of agent i over agent j is equal to j's influence over i. Therefore, $\boldsymbol{A} = \boldsymbol{A}^T$ and the undirected DeGroot model is a symmetric model. In the directed case, we make no assumption on the relationship of w_{ij} and w_{ji}, therefore the directed DeGroot model is an asymmetric model and thus provides richer dynamics of opinion formation and allows the analysis of more complex agent behavior.

Given a vector $\boldsymbol{x}(0)$ of initial agent opinions, the vector of the agents' opinions at time t is given by

$$\boldsymbol{x}(t) = \boldsymbol{A}^t\boldsymbol{x}(0) \tag{2.3}$$

Perhaps the most interesting question in the field of opinion dynamics is whether the opinions in a particular model converge to a limit in the long run and, if so, how fast the vector of opinions approaches that limit. Standard results in Markov chain theory can be applied in the DeGroot model, since $\boldsymbol{A}$ is stochastic, to show that it converges to a stable state, under certain assumptions.

Theorem 2.1. *The DeGroot model with n agents converges to the unique equilibrium point*

$$\boldsymbol{x}^* = \lim_{t\to\infty} \boldsymbol{x}(t) = \lim_{t\to\infty} \boldsymbol{A}^t \boldsymbol{x}(0) \tag{2.4}$$

for any initial vector of opinions $\boldsymbol{x}(0) \in [0,1]^n$*, if and only if the Markov chain with transition matrix* $\boldsymbol{A}$ *is irreducible and aperiodic.*

In our setting these two condition imply that for every agent i, there exists a time t_0 such that for every time $t \geq t_0$, i is influenced (albeit indirectly) by all other agents. This is equivalent to the matrix $\boldsymbol{A}^{t_0}$ having only positive elements. Since

$$\boldsymbol{x}^* = \boldsymbol{A}\boldsymbol{x}^* \tag{2.5}$$

we will call the limit vector $\boldsymbol{x}^*$, the *Nash equilibrium* of the model.

The update rule in the DeGroot model is illustrated in the following examples.

Example 2.1 (The DeGroot Model)**.** *Consider an instance of the DeGroot model, with* $n = 3$ *agents, and a transition matrix*

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{2.6}$$

In this example, the first agent weighs the opinions of the two other agents equally, the second agent listens only to the first and the third agent listens only to the second.

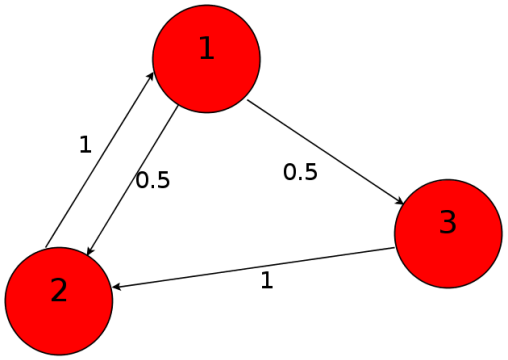

FIGURE 2.1: An example of a convergent DeGroot model

If we consider a vector of initial opinions

$$\boldsymbol{x}(0) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \tag{2.7}$$

then applying (2.2), to get the opinions at time $t = 1$*, gives us*

$$\boldsymbol{x}(1) = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \\ 1 \end{bmatrix}$$

Then, the agents update their opinions again, and we get

$$\boldsymbol{x}(2) = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$$

Note that the above Markov chain is irreducible and we can also show that it is aperiodic. Hence, from Theorem 2.1, it converges to the unique Nash equilibrium.

$$\boldsymbol{x}^* = \lim_{t\to\infty} \boldsymbol{A}^t \boldsymbol{x}(0) = \begin{bmatrix} 2/5 & 2/5 & 1/5 \\ 2/5 & 2/5 & 1/5 \\ 2/5 & 2/5 & 1/5 \end{bmatrix} \boldsymbol{x}(0) \tag{2.8}$$

Therefore, we see that $\boldsymbol{x}^* = \begin{bmatrix} 2/5 & 2/5 & 2/5 \end{bmatrix}^T$*, and all agents have converged to the same opinion.*

Example 2.2 (The DeGroot Model - Non-convergence)**.** *Consider the previous example, slightly alterated so that the third agent listens only to the first agent. Then, this instance of the DeGroot model is not aperiodic, but has a period of* 2 *instead. Thus, it violates the premises of Theorem 2.1, and we can show that the agents' opinions never converge in this case. Indeed, the transition matrix is*

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{2.9}$$

In this example, the first agent again weighs the opinions of the two other agents equally, but the other two agents listen only to the first.

We consider the same vector of initial opinions as in the previous example

$$\boldsymbol{x}(0) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \tag{2.10}$$

Since the period of our instance is 2*, we can calculate the transition matrix for even and odd time steps. For any* $k \geq 1$*, we have*

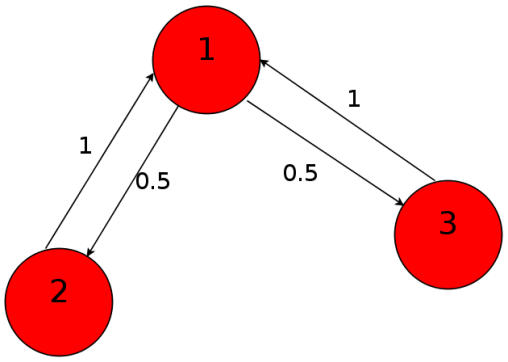

FIGURE 2.2: An example of a non-convergent DeGroot model

$$\boldsymbol{A}^{2k} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix} \text{ and } \boldsymbol{A}^{2k+1} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{2.11}$$

We observe that the transition matrix alternates between two matrices, at even and odd times. Thus, the limit $\lim_{t\to\infty} \boldsymbol{A}^t$ *does not exist, and the agents' opinions do not converge. Intuitively, the agents interchange their opinions at each time step.*

A more careful analysis of the DeGroot model will provide information about the Nash equilibrium as well as the rate of convergence to it. The Perron-Frobenius theorem [21] states that, since $\boldsymbol{A}$ is stochastic, its *spectral radius* $\rho(\boldsymbol{A})$, i.e. its largest eigenvalue, is equal to 1. Also, since there exists a t_0 such that $\boldsymbol{A}^{t_0}$ has only positive elements, $\rho(\boldsymbol{A})$ is a unique eigenvalue, and it corresponds to the eigenvector $\boldsymbol{q}_1 = \frac{1}{n}\mathbf{1} = \left[\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right]$. Let the eigenvalues of $\boldsymbol{A}$ be enumerated in decreasing order of their absolute values, such that $1 = |\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|$, and let $\boldsymbol{Q}$ be the matrix of eigenvectors of $\boldsymbol{A}$, where each column of $\boldsymbol{Q}$ is a eigenvector of $\boldsymbol{A}$, normalized to having an L_2 norm equal to 1. Furthermore, let $\boldsymbol{\Lambda}$ be the diagonal matrix of the eigenvalues of $\boldsymbol{A}$, i.e. $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Since the eigenvectors of $\boldsymbol{A}$ are orthonormal, they are linearly independent, and thus $\boldsymbol{A}$ can be factorized as

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1} \tag{2.12}$$

which gives

$$\boldsymbol{A}^t = \boldsymbol{Q}\boldsymbol{\Lambda}^t\boldsymbol{Q}^{-1} \tag{2.13}$$

Since the eigenvectors $\boldsymbol{q}_i$ of $\boldsymbol{A}$ are linearly independent, they span $\mathbb{R}^n$, and we can write

$$\boldsymbol{x}^T(0) = \sum_{i=1}^{n} c_i \boldsymbol{q}_i \tag{2.14}$$

for some set of $c_i \in \mathbb{R}$. Therefore

$$\boldsymbol{x}(t) = \boldsymbol{A}^t \boldsymbol{x}(0) = \sum_{i=1}^{n} c_i \lambda_i^t \boldsymbol{q}_i = c_1 \boldsymbol{q}_1 + \sum_{i=2}^{n} c_i \lambda_i^t \boldsymbol{q}_i \tag{2.15}$$

From the equation above, it is easy to observe that when $t \to \infty$, $\boldsymbol{x}^* = c_1 \boldsymbol{q}_1$. Therefore, the unique Nash equilibrium of the DeGroot model is a state in which all agents reach consensus. In addition, since λ_2 dominates all other eigenvalues in the order of their absolute values, except for $\lambda_1 = 1$, the rate of convergence to $\boldsymbol{x}^*$ is exponential in the order of λ_2.

2.1.2 The Friedkin - Johnsen Model

The DeGroot model, while deceptively simple, can be used to model any linear behavior of the agents. However, it is interesting to develop extensions of the DeGroot model that simulate real-world social netowrks in a better fashion. One such extension is the *Friedkin-Johnsen (FJ) model*, introduced by Friedkin and Johnsen in 1990 [4]. In the FJ model, each agent i, apart from her expressed opinion x_i, holds a persistent intrinsic opinion s_i. This internal opinion remains constant even as agent i updates her overall opinion $x_i(t)$ through averaging. At each time step t, each agent i updates her expressed opinion to

$$x_i(t+1) = w_{ii} s_i + \sum_{\substack{j \in \mathcal{N}_i \\ j \neq i}} w_{ij} x_j(t) \tag{2.16}$$

or, in matrix form

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{s} \tag{2.17}$$

where the elements in the diagonal of $\boldsymbol{A}$ are equal to 0, $\boldsymbol{B}$ is a diagonal matrix with the element at position (i, i) being equal to w_{ii}, and $\boldsymbol{s}$ is the vector of the agents' intrinsic opinions. As with the DeGroot model, we can assume, without loss of generality, that $\sum_{j \in \mathcal{N}_i} w_{ij} = 1$. We require at least one $w_{ii} \neq 0$ for some agent i, so that we do not have an instance of the DeGroot model. If we make the logical assumption that the vector of initial agent opinions $\boldsymbol{x}(0) = \boldsymbol{s}$, then iterating (2.17) shows that the vector of opinions at each time $t \geq 0$ is

$$\boldsymbol{x}(t) = \boldsymbol{A}^t \boldsymbol{s} + \sum_{k=0}^{t-1} \boldsymbol{A}^k \boldsymbol{B} \boldsymbol{s} \tag{2.18}$$

It is easy to see that the FJ model can be simulated via the DeGroot model. Indeed, if we consider an instance of the FJ model, we can set $w_{ii} = 0$, and add, for each agent i, an imaginary new agent g_i in G, with the following properties

- $x_{g_i}(t) = s_i$, for every agent g_i and any time t.
- $w_{g_i g_i} = 1$.
- $w_{g_i k} = 0$, for every agent $k \neq g_i$.
- $w_{j g_i} = 0$, for every agent $j \neq i$.
- $w_{i g_i} = w_{ii}$, for every agent i.

However, the FJ model differs from the DeGroot model in the sense that $\boldsymbol{A}$ is now a substochastic matrix. Therefore, standard Markov chain theory results can be applied again to show that $\rho(\boldsymbol{A}) < 1$. If we limit our analysis of this model to the undirected case, where $w_{ij} = w_{ji}$ for every pair of agents i and j, we can show that the undirected FJ model admits to a Nash equilibrium.

Theorem 2.2. *Consider an instance of the undirected FJ model with n agents. Then, the opinions of the agents converge to the unique Nash equilibrium*

$$\boldsymbol{x}^* = \sum_{k=0}^{\infty} \boldsymbol{A}^k \boldsymbol{B} \boldsymbol{s} = (\boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{B} \boldsymbol{s} \tag{2.19}$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix.

Additionally, if we denote an upper bound γ on the distance from the equilibrium, at time t, as $\|\boldsymbol{x}(t) - \boldsymbol{x}^*\|_\infty \leq \gamma$, we can show that the dynamics of (2.17) converge, for the undirected case, to $\boldsymbol{x}^*$ in $\mathcal{O}\big(\frac{ln(n/\gamma)}{1-\rho(\boldsymbol{A})}\big)$ time steps [22].

The update rule in the FJ model will become clearer in the following example.

Example 2.3 (The Friedkin-Johnsen Model)**.** *Consider an instance of the FJ model, with $n = 3$ agents, and a transition matrix*

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1/4 & 1/3 \\ 1/4 & 0 & 1/2 \\ 1/3 & 1/2 & 0 \end{bmatrix} \tag{2.20}$$

We also have a vector of intrinsic opinions $s = \begin{bmatrix} 1 & 1/2 & 3/4 \end{bmatrix}^T$*, along with a diagonal matrix with the weight that each agent has for his intrinsic opinion*

$$\boldsymbol{B} = \begin{bmatrix} 5/12 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/6 \end{bmatrix} \tag{2.21}$$

We consider the vector of initial opinions $\boldsymbol{x}(0) = s$*, and we apply (2.17) to get the opinions at time* $t = 1$

$$\begin{aligned} \boldsymbol{x}(1) &= \begin{bmatrix} 0 & 1/4 & 1/3 \\ 1/4 & 0 & 1/2 \\ 1/3 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1/2 \\ 3/4 \end{bmatrix} + \begin{bmatrix} 5/12 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/6 \end{bmatrix} \begin{bmatrix} 1 \\ 1/2 \\ 3/4 \end{bmatrix} \\ &= \begin{bmatrix} 3/8 \\ 5/8 \\ 7/12 \end{bmatrix} + \begin{bmatrix} 5/12 \\ 1/8 \\ 1/8 \end{bmatrix} = \begin{bmatrix} 19/24 \\ 18/24 \\ 17/24 \end{bmatrix} \end{aligned}$$

From Theorem 2.2 we have that this undirected FJ model converges to the unique Nash equilibrium

$$\boldsymbol{x}^* = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B}\boldsymbol{s} = \begin{bmatrix} \frac{243}{284} \\ \frac{103}{142} \\ \frac{439}{568} \end{bmatrix} \tag{2.22}$$

It is important to note that, since $B_{ii} = 0$ for all non-stubborn agents by definition, the Nash equilibrium $\boldsymbol{x}^*$ depends only on the initial opinions of the stubborn agents, and the initial opinions of the non-stubborn agents eventually vanish and have no effect on the equilibrium. Furthermore, the presence of stubborn agents indicate that the agents do not reach a consensus, but the dynamics of (2.17) converge to an equilibrium in which the opinion of each agent is a convex combination of the initial opinions of the stubborn agents.

2.2 Non-Linear (Coevolutionary) Models

We will continue our presentation of the main models used in the field of opinion dynamics with non-linear models. In contrast with the previous models, non-linear models allow the graph G_t to change over time. More specifically, the neighborhood $\mathcal{N}_i$ of agent i changes at every time t, as each agent chooses to be influenced by a

different subset of agents each time. Thus, $E(G_t)$ changes over time and edges are added, deleted, or have their weights adjusted, to represent the fluctuation of $\mathcal{N}_i$. Non-linear models are also called *coevolutionary* models, since the opinions of the agents and the underlying graph of the social network coevolve, with one affecting the other. Obviously, coevolutionary models can simulate a wider range of real-world phenomena, since agents in real social networks constantly update who they are influenced by, and allow for vastly richer dynamics.

The main difference in our analysis of coevolutionary models is that powerful linear techniques such as matrix theory, Markov chains and graph theory, used to analyze linear models, are no longer applicable. More specifically, while most of our notation remains the same from our analysis of the linear models, it is understood that, as one would expect, $\boldsymbol{A}(t)$ is a time-variant matrix in coevolutionary models. This fact makes the analysis of such interesting models, and the computation of rigorous analytical results for them, considerably more difficult than the previous ones.

2.2.1 The Hegselmann - Krause Model

Perhaps the most studied and well-known coevolutionary model, due to its generality and ability to capture almost every notion of agent interaction either directly or via one of its variants, is the *Hegselmann-Krause (HK) model*, first introduced in 2002 by Hegselmann and Krause [5]. In the HK model, we are given a vector of initial opinions $\boldsymbol{x}(0)$ of the n agents, along with their confidence $\varepsilon > 0$. The confidence of the agents is used in their computation of their opinion-dependent neighborhood, and is a constant value, uniform for all agents, that characterizes each instance of the HK model. Since ε characterizes the neighborhood of every agent, the HK model is also known as the *bounded confidence model*. At each time $t \geq 1$, every agent i computes her neighborhood

$$\mathcal{N}_i(t, \varepsilon) = \{j : |x_i(t-1) - x_j(t-1)| \leq \varepsilon\} \tag{2.23}$$

and updates her opinion to the average of the opinions in $\mathcal{N}_i(t, \varepsilon)$

$$x_i(t) = \sum_{j \in \mathcal{N}_i(t,\varepsilon)} \frac{x_j(t-1)}{|\mathcal{N}_i(t, \varepsilon)|} \tag{2.24}$$

or, in matrix form

$$\boldsymbol{x}(t) = \boldsymbol{A}(t, \boldsymbol{x}(t))\boldsymbol{x}(t-1) \tag{2.25}$$

In the HK model, each $\boldsymbol{A}(t, \boldsymbol{x}(t))$ is a $n \times n$ matrix with the (i, j) element being $1/|\mathcal{N}_i(t, \varepsilon)|$ if $j \in \mathcal{N}_i(t, \varepsilon)$ and 0 otherwise. Therefore, $\boldsymbol{A}(t, \boldsymbol{x}(t))$ corresponds to the adjacency matrix of G_t at every time t. For convenience, in the rest of this thesis, we will denote $\boldsymbol{A}(t, \boldsymbol{x}(t))$ by $\boldsymbol{A}_t$. Iterating (2.25), we get

$$\boldsymbol{x}(t) = \boldsymbol{A}_t \boldsymbol{A}_{t-1} \dots \boldsymbol{A}_1 \boldsymbol{x}(0) \tag{2.26}$$

which shows that the opinions of the agents at time t is equal to the application of a series of different linear transformations on the vector of initial agent opinions, with each such transformation being dependent on the neighborhood of every agent.

Next, we present an example that illustrated the update rule in the HK model.

Example 2.4 (The Hegselmann-Krause Model)**.** *Consider an instance of the HK model, with $n = 3$ agents, where $\varepsilon = \frac{1}{2}$ and the initial opinions of the agents are*

$$\boldsymbol{x}(0) = \begin{bmatrix} 0 & 1/2 & 1 \end{bmatrix}^T \tag{2.27}$$

Each agent computes her neighborhood at time 1

$$\begin{aligned} \mathcal{N}_1(1, \varepsilon) &= \{1, 2\} \\ \mathcal{N}_2(1, \varepsilon) &= \{1, 2, 3\} \\ \mathcal{N}_3(1, \varepsilon) &= \{2, 3\} \end{aligned}$$

and then updates her opinion to the average of the opinions in her neighborhood

$$\boldsymbol{x}(1) = \begin{bmatrix} 1/4 & 1/2 & 3/4 \end{bmatrix}^T \tag{2.28}$$

Then, we repeat the process for $t = 2$. Note that, now, the agents' neighborhoods have changed

$$\begin{aligned} \mathcal{N}_1(2, \varepsilon) &= \{1, 2, 3\} \\ \mathcal{N}_2(2, \varepsilon) &= \{1, 2, 3\} \\ \mathcal{N}_3(2, \varepsilon) &= \{1, 2, 3\} \end{aligned}$$

Again, each agent updates her opinion to the average of the opinions in her neighborhood

$$\boldsymbol{x}(2) = \begin{bmatrix} 1/2 & 1/2 & 1/2 \end{bmatrix}^T \tag{2.29}$$

We see that the agents reached consensus very quickly, with every agent converging to $x^* = \frac{1}{2}$*. Next, we investigate whether converge is guaranteed in the HK model or whether it was an idiosyncracy of this particular example.*

Before we continue investigating the convergence properties of the HK model, we should note some important properties of the model. We first define the concept of a *split* between two agents

Definition 2.3 (Split)**.** Consider two agents i and j. If we have $|x_i(t-1) - x_j(t-1)| \leq \varepsilon$ at time $t-1$ but $|x_i(t) - x_j(t)| > \varepsilon$ at time t, we call this event a *split* at time t, because this leads to $i \notin \mathcal{N}_j(t+1, \varepsilon)$ and $j \notin \mathcal{N}_i(t+1, \varepsilon)$.

We continue with some interesting properties of the HK model.

- The dynamics of (2.25) do not change the order of the agents' opinions. Specifically, if for two agents i and j we have $x_i(t) \leq x_j(t)$, this implies that $x_i(t+1) \leq x_j(t+1)$.
- If a split between two agents occurs at time t_0, this implies that the split between these agents will remain for all times $t \geq t_0$. Thus, the agents behave independently and do not affect each other after t_0. However, this property is not true in higher dimensions, i.e. when $x_i \in \mathbb{R}^d$.
- Finally, if for a specific instance of initial agent opinions $\boldsymbol{x}(0)$ the HK model converges to a consensus, this implies that $|x_i(t) - x_j(t)| \leq \varepsilon$ for all agents i, j at all times $t \geq 0$.

There is a considerable amount of research focused on the convergence properties of the HK model. There are various results that prove convergence of the HK model under certain assumptions [23–25]. Bhattacharyya et al provided interesting upper and lower bounds on the convergence rate of the HK model [26]. Specifically, they proved that the 1-dimensional HK model converges in $\mathcal{O}(n^3)$ time and the d-dimensional HK model converges in $poly(n, d)$ time. They also provided a lower bound of $\Omega(n^2)$ in the 1-dimensional case. The intuition behind their proof of $\mathcal{O}(n^3)$ is particularly interesting. Essentially, they prove that for a group of agents where no split occurs, the time required for them to reach convergence is $\mathcal{O}(n^2)$. Since at most n splits can occur, the 1-dimensional HK model converges in $\mathcal{O}(n^3)$ time. Furthermore, we also know that the instance of HK model where the agents' opinions lie on a circle instead of a line also converges [27].

In the following chapters we are going to focus on the HK model, introduce variants and extensions, present certain results on their convergence properties along with the most important mathematical tools used in the analysis of non-linear models.

2.2.2 The Deffuant - Weisbuch Model

Up until now, all the models we have presented, linear and non-linear, share a common property in that they are deterministic. The last model we will present here differs from that scope as it introduces randomness in the process through which the agents' opinions are updated. In the *Deffuant - Weisbuch (DW) model*, introduced by Deffuant and Weisbuch [6, 7], we consider n agents with an initial vector of opinions $\boldsymbol{x}(0) \in [0,1]^n$. At each time step t, two randomly chosen agents meet and re-adjust their opinions if and only if their difference in opinion is smaller in magnitude than a certain threshold confidence ε. We also consider a convergence parameter $\mu \in [0, \frac{1}{2}]$, and if at time $t \geq 1$ agents i and j are chosen to meet, they update their opinions to

$$x_i(t) = \begin{cases} x_i(t-1), & if\ |x_i(t-1) - x_j(t-1)| > \varepsilon \\ (1-\mu)\, x_i(t-1) + \mu\, x_j(t-1), & if\ |x_i(t-1) - x_j(t-1)| \leq \varepsilon \end{cases} \tag{2.30}$$

$$x_j(t) = \begin{cases} x_j(t-1), & if\ |x_i(t-1) - x_j(t-1)| > \varepsilon \\ (1-\mu)\, x_j(t-1) + \mu\, x_i(t-1), & if\ |x_i(t-1) - x_j(t-1)| \leq \varepsilon \end{cases} \tag{2.31}$$

Therefore, their updated opinions are a convex combination of their old opinions. In the DW model, ε is considered constant both in time and across all the agents.

The exchange of opinions in the DW model can also be represented in matrix form

$$\boldsymbol{x}(t) = \begin{cases} \boldsymbol{x}(t-1), & if\ |x_i(t-1) - x_j(t-1)| > \varepsilon \\ \boldsymbol{A}_{ij}\boldsymbol{x}(t-1), & if\ |x_i(t-1) - x_j(t-1)| \leq \varepsilon \end{cases} \tag{2.32}$$

where $\boldsymbol{A}_{ij}$ is an $n \times n$ matrix equal to the identity matrix $\boldsymbol{I}$ except for the elements $a_{ii} = a_{jj} = 1 - \mu$ and $a_{ij} = a_{ji} = \mu$. It is easy to see that $\boldsymbol{A}_{ij} = \boldsymbol{A}_{ij}^T$, thus the DW model is a symmetric model.

The distinct update rule of the DW model can be properly explained with an example.

Example 2.5 (The Deffuant-Weisbuch Model)**.** *Consider an instance of the DW model, with $n = 3$ agents, where $\varepsilon = \frac{2}{3}$, $\mu = \frac{1}{3}$ and the initial opinions of the agents are*

$$\boldsymbol{x}(0) = \begin{bmatrix} 0 & 1/2 & 1 \end{bmatrix}^T \tag{2.33}$$

Let an external entity randomly choose two agents for $t = 1$*, say agents* 1 *and* 2*. The random selection follows the uniform distribution over all agents. Then, agents* 1 *and* 2 *update their opinions, since their difference is smaller than* ε*, while agent* 3*'s opinion stays the same.*

$$\boldsymbol{x}(1) = \begin{bmatrix} 1/6 & 1/3 & 1 \end{bmatrix}^T \tag{2.34}$$

Next, let 1 *and* 3 *be the chosen agents, and repeat the process for* $t = 2$*. However, their difference is larger than* ε*, therefore no agent changes her opinion at time* $t = 2$*.*

$$\boldsymbol{x}(2) = \boldsymbol{x}(1) = \begin{bmatrix} 1/6 & 1/3 & 1 \end{bmatrix}^T \tag{2.35}$$

Finally, if agents 2 *and* 3 *are chosen at time* $t = 3$*, we get*

$$\boldsymbol{x}(3) = \begin{bmatrix} 1/6 & 5/9 & 7/9 \end{bmatrix}^T \tag{2.36}$$

It is important to note that the DW model differs in many ways from the previous models. As stated before, it is an inherently random model, in contrast to the previous models which were deterministic. In addition, while in all previous models all agents updated their opinions simultaneously at each time step, the DW is a serial model, since at each time t only two agents, say i and j, interact and possibly update their opinions. The remaining agents $k \neq i, j$ do not update their opinions at time t. This means that, while highly unlikely, it is possible for an agent k to not be chosen for all times t up to a fixed time t_0, therefore having $x_k(t) = x_k(0)\ \forall t \leq t_0$. In this case, agent k behaves as a fully-stubborn agent that never updates her opinion.

The DW model is known to converge to an equilibrium $\boldsymbol{x}^*$ [24, 28]. While no significant upper or lower bound on the convergence time is known, the DW model is known to converge exponentially fast, even in the asymmetric case, under certain assumptions [29].

Chapter 3

Variations of the Hegselmann - Krause model

In this chapter, we focus our attention on the Hegselmann - Krause model. Although the HK model is considered the basis of all non-linear models, thus making it perhaps the most important one, and it is definitely an improvement from basic, linear models, it is very simple in its definition and it cannot properly capture the complex interactions of real-world social networks. Therefore, developing and studying variations of the HK model has the potential of producing theoretical results that have greater significance for real social networks and can be more easily applied. We introduce several interesting variations of the HK model and present certain important results on their convergence properties.

3.1 The Network-HK Model

A natural extension of the HK model, and the first variation we will present in this chapter, is the *Network-HK model* [30]. It extends the HK model with the addition of an underlying social network that limits the possible ways opinions are shared between the agents. Therefore, it introduces the concept of opinion update under limited information.

3.1.1 Definition

In the Network-HK model, along with the vector of initial opinions $\boldsymbol{x}(0)$ and the agents' confidence ε, we are also given a undirected weighted connected graph $G(V, E)$, that represents an underlying social network. The graph may change over time, but it follows a specific set of rules. In particular, G and its adjacency matrix $\boldsymbol{A}$ must satisfy the following properties

1. For every agent i, $w_{ii} > 0$ at all times $t \geq 0$. This implies that all agents have self-loops in G and the diagonal of $\boldsymbol{A_t}$ is strictly positive.

2. For every pair of agents i and j, $w_{ij} > 0 \iff w_{ji} > 0$. This implies that G is bidirectional, in the sense that confidence is mutual for all agents.

3. If we denote the minimum positive weight at time t by $w^*(t) \coloneqq min_{w_{ij}>0}\, w_{ij}$, there exists a fixed $\delta > 0$, such that $w^*(t) > \delta$ at all times $t \geq 0$. This implies that positive weights in $\boldsymbol{A}_t$ do not converge to zero.

Any graph G given in the Network-HK model must satisfy these properties and the results presented here hold for any such G. In this model, at every time step t, each agent i computes her neighborhood

$$\mathcal{N}_i(G_t, t, \varepsilon) = \{j : \{i, j\} \in E \; and \; |x_i(t-1) - x_j(t-1)| \leq \varepsilon\} \tag{3.1}$$

and updates her opinion to the average of the opinions in $\mathcal{N}_i(G_t, t, \varepsilon)$, as in the HK model. Note that (2.26) holds in the Network-HK model as well, albeit with some modifications on the definition of $\boldsymbol{A}_t$. Specifically, each $\boldsymbol{A}_t$ is an $n \times n$ matrix with the (i, j) element being $1/|\mathcal{N}_i(G_t, t, \varepsilon)|$ if $j \in \mathcal{N}_i(G_t, t, \varepsilon)$ and 0 otherwise. Furthermore, since $w_{ii} > 0$ for any agent i and any time t, it is understood that $i \in \mathcal{N}_i(G_t, t, \varepsilon)$. Also note that $\boldsymbol{A}_t$ corresponds to the adjacency matrix of an undirected, unweighted graph, but with normalized rows so that each $\boldsymbol{A}_t$ is stochastic.

3.1.2 Results

Fotakis et al [30], who introduced the Network-HK model, also provide a proof of its convergence to a stable state. They observe that, if at some point the underlying graph G is disconnected, the agents are partitioned into two subsets $(S, V \setminus S)$ such that no edge $\{i, j\}$ between agents $i \in S$ and $j \in V \setminus S$ traverses the cut $(S, V \setminus S)$. Therefore, there is no influence between the agents in S and in $V \setminus S$.

Now, assume there exists a time t_0 such that no edges traverse the cut $(S, V \setminus S)$ for all times $t \geq t_0$. It immediately follows that, after t_0, no agent in S can ever again influence an agent in $V \setminus S$, and vice-versa. Thus, at t_0, our system *breaks* into independent subsystems. Intuitively, since at most $|V| - 1$ breaks occur, there exists a finite time t^* after which no breaks occur. Therefore, if we can prove convergence for the Network-HK model, given that no breaks occur, this would imply convergence for each of our subsystems. Therefore, the overall Network-HK model will converge as well.

We start by defining the concept of a *weakly connected set* of agents, that properly captures the concept of a set of agents where no breaks occur, for any time t.

Definition 3.1 (Weakly Connected Set)**.** We say that a set of agents $S \subseteq V$ is *weakly connected* if for any $S' \subset S$ such that $S' \neq \emptyset$ and any $t_0 \in \mathbb{N}$, there is a round $t \geq t_0$ so that G_t includes at least one edge connecting an agent in S' to some agent in $S \setminus S'$.

Our definition of weak connectivity attempts to capture the notion that all agents in a weakly connected set influence each other with their opinions. Observe that it is equivalent to the negaiton of the property that a break occurs at some time step t_0. If the set of agents V is not weakly connected, it can be uniquely partitioned into weakly connected components, as the following lemma demonstrates.

Lemma 3.2. *Given a graph $G(V, E)$, there is a unique partition of V into weakly connected components $V_1, V_2, \ldots, V_m$.*

Proof. Consider an agent $i \in V$ and let V_1 be a maximal weakly connected set that includes i. We assume that for any time step t_1 there exists a time step $t_2 \geq t_1$ such that there exists an edge in G_{t_2} that connects an agent in V_1 to an agent in $V \setminus V_1$. Since $V \setminus V_1$ contains a finite number of agents, this implies the existence of an agent $j \in V \setminus V_1$ such that for any time step t_1 there exists a time step $t_2 \geq t_1$ where G_{t_2} contains an edge connecting an agent in V_1 to j. This means that $V_1 \cup \{j\}$ is also weakly connected, which contradicts the maximality of V_1. Therefore, our claim was wrong and there exists a time step t_0 such that for all time steps $t \geq t_0$, there is no edge in G_t that connects any agent in V_1 to any agent in $V \setminus V_1$.

Consider now an agent $i' \in V \setminus V_1$. Following the previous thought process, let V_2 be a maximal weakly connected set that includes i'. It is easy to observe that $V_1 \cap V_2 = \emptyset$, since, if we consider an agent $j' \in V_1 \cap V_2$, we get that $V_1 \cup \{j'\}$ is also weakly connected, which again contradicts the maximality of V_1. Therefore, V_1 is the unique maximal weakly connected set that includes i. We continue inductively, applying the same argument to $V \setminus V_1$, to obtain a unique partition of V into weakly connected components $V_1, V_2, \ldots, V_m$. □

From Lemma 3.2, it is understood that we can focus our analysis of the Network-HK model in the case where V is a weakly connected set of agents. Indeed, utilizing the notion of the *coefficient of ergodicity* of a matrix, we can prove that the agents of a weakly connected set reach consensus. First, we define the coefficient of ergodicity of a stochastic, square matrix.

Definition 3.3 (Coefficient of Ergodicity)**.** The *coefficient of ergodicity* of a stochastic, $n \times n$ matrix $\boldsymbol{A}$, denoted by $\tau(\boldsymbol{A})$, is defined as $\tau(\boldsymbol{A}) \coloneqq \frac{1}{2}\, max_{i,j} \|\boldsymbol{A}^T(\boldsymbol{e}_i - \boldsymbol{e}_j)\|_1$, where $\boldsymbol{e}_i$ is the vector with 1 in coordinate i and 0 in all other coordinates.

Before we proceed, we will present some important properties of the coefficient of ergodicity. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be stochastic matrices. Then

i. $\tau(\boldsymbol{A}) \leq 1$.

ii. $\tau(\boldsymbol{AB}) \leq \tau(\boldsymbol{A})\tau(\boldsymbol{B})$.

iii. $\tau(\boldsymbol{A}) = 0 \iff rank(\boldsymbol{A}) = 1$.

iv. $\forall i, j \ a_{ij} > 0 \implies \tau(\boldsymbol{A}) < 1$.

Now that we have defined the necessary concepts needed for our proof, we can proceed with the following lemma, which proves that a weakly connected set of agents reaches consensus.

Lemma 3.4. *Let $(G(V, E), \varepsilon, \boldsymbol{x}(0))$ be an instance of the Network-HK model, where V is weakly connected. Then, all agents converge to a single opinion x^*.*

Proof. To prove the lemma, we will show that there exists a time step $t_0 \geq 0$, such that the matrix $\boldsymbol{C}_1^{t_0} = \boldsymbol{A}_{t_0}\boldsymbol{A}_{t_0-1}\dots\boldsymbol{A}_1$ has $\tau(\boldsymbol{C}_1^{t_0}) \leq \varepsilon/2$. We also have, from (2.26) that $\boldsymbol{x}(t_0) = \boldsymbol{C}_1^{t_0}\boldsymbol{x}(0)$. Combining the two equations above, we get that for all agents i and j

$$|x_i(t_0) - x_j(t_0)| = |(\boldsymbol{e}_i\boldsymbol{C}_1^{t_0} - \boldsymbol{e}_j\boldsymbol{C}_1^{t_0})\boldsymbol{x}(0)| \leq \|\boldsymbol{e}_i\boldsymbol{C}_1^{t_0} - \boldsymbol{e}_j\boldsymbol{C}_1^{t_0}\|_1 \tag{3.2}$$

where $\boldsymbol{e}_i\boldsymbol{C}_1^{t_0}$ is equal to the i-th row of matrix $\boldsymbol{C}_1^{t_0}$, and the last inequality holds due to $\boldsymbol{x}(0) \in [0, 1]^n$. Therefore we have

$$|x_i(t_0) - x_j(t_0)| \leq \|\boldsymbol{e}_i\boldsymbol{C}_1^{t_0} - \boldsymbol{e}_j\boldsymbol{C}_1^{t_0}\|_1 \leq 2\tau(\boldsymbol{C}_1^{t_0}) \leq \varepsilon \tag{3.3}$$

We see that at time t_0, all agents are within distance ε, thus at any time $t \geq t_0$, all agents compute the average of all the opinions in their social neighborhood, which includes their opinion. This implies that $\boldsymbol{A}_t = \boldsymbol{A}_{t_0}$ for any time $t \geq t_0$, which means that $\boldsymbol{A}$ is a constant matrix after t_0. We can easily see that this is essentialy an instance of the undirected DeGroot model, where $w_{ii} > 0$ for any agent i. From V's weak connectivity, we get that the process defined by $\boldsymbol{A}_{t_0}$ is irreducible and, since $w_{ii} > 0$ for any agent i, we get that the process is aperiodic. Therefore, by [1], all agents converge to a single opinion x^*.

It remains to prove for any $\delta > 0$ the existence of a time step $t_0 \geq 0$, such that the matrix $\boldsymbol{C}_1^{t_0} = \boldsymbol{A}_{t_0}\boldsymbol{A}_{t_0-1}\dots\boldsymbol{A}_1$ has $\tau(\boldsymbol{C}_1^{t_0}) \leq \delta$. Then, setting $\delta = \varepsilon/2$, we get the necessary requirements for (3.3) to hold.

First of all, we prove that the weak connectivity of V implies that for any time step t, there exists a corresponding time step $l(t) \geq t$, such that the matrix $\boldsymbol{C}_t^{l(t)} = \boldsymbol{A}_{l(t)}\boldsymbol{A}_{l(t)-1}\dots\boldsymbol{A}_t$ has only positive elements. Thus, from property (iv) of the coefficient of ergodicity, we get $\tau(\boldsymbol{C}_t^{l(t)}) < 1$. To prove this claim, we note that the (i,j) element of $\boldsymbol{C}_t^{l(t)}$ is positive if and only if there is a time-respecting walk $(q_0 = i, q_1, \dots, q_{m-1}, q_m = j)$ from agent i to agent j such that

- Edge $\{i, q_1\}$ exists in $G_{t'}$ for some $t' \geq t$.
- For every index $1 \leq k \leq m-1$, if the edge $\{q_{k-1}, q_k\}$ exists in $G_{t'}$ for some t', then the edge $\{q_k, q_{k+1}\}$ exists in $G_{t''}$ for some $t'' > t'$.

Since each $\boldsymbol{A}_t$ has positive diagonal elements, the walk can wait at each intermediate agent until the next edge appears. The existence of such a walk between all pairs of agents in V follows directly from the definition of weak connectivity.

We now know that for any time t there is a corresponding time $l(t)$ such that $\tau(\boldsymbol{C}_t^{l(t)}) < 1$. However, this is not enough to prove our lemma. We also need to show that there is a fixed $\eta > 0$ such that $\tau(\boldsymbol{C}_t^{l(t)}) \leq 1 - \eta$. Then, concatenating p non-overlapping sequences $\boldsymbol{A}_t, \dots, \boldsymbol{A}_{l(t)}$ and utilizing property (ii) of the coefficient of ergodicity, we get a sequence $\boldsymbol{A}_1, \dots, \boldsymbol{A}_{t_0}$ such that $\tau(\boldsymbol{A}_1 \dots \boldsymbol{A}_{t_0}) \leq (1-\eta)^p$. For an appropriately large p, we get $\tau(\boldsymbol{A}_1 \dots \boldsymbol{A}_{t_0}) \leq \delta$ for any $\delta > 0$.

To show that $\tau(\boldsymbol{C}_t^{l(t)}) \leq 1 - \eta$, we note that there exist at most n^2 matrices in the sequence $\boldsymbol{A}_t, \dots, \boldsymbol{A}_{l(t)}$ that strictly increase the number of agents reachable from some other agent by time-respecting walks, equal to the number of all possible pairs of n agents in G. Therefore, $\tau(\boldsymbol{A}_t, \dots, \boldsymbol{A}_{l(t)})$ is bounded by the product of the coefficients of ergodicity of these specific matrices. Since the number of these matrices is at most n^2 and the number of different matrices $\boldsymbol{A}_t$ is finite and equal to the number of induced subgraphs of G, i.e. all the possible subsets of V along with their respective edges, there exists a fixed $\eta > 0$ such that $\tau(\boldsymbol{A}_t, \dots, \boldsymbol{A}_{l(t)}) \leq 1 - \eta$ and our proof of Lemma 3.4 is complete. □

Intuitively, a weakly connected set of agents reaches consensus because all agents influence (directly or indirectly) each other with their opinions, for all times t. Therefore, the maximum distance between the agents' opinions decreases as time goes by, leading all agents to converge to the same opinion. It should be obvious that Lemmas 3.2 and 3.4 guarantee the convergence of the Network-HK model.

Theorem 3.5. *The Network-HK model converges to a stable state.*

Proof. From Lemma 3.2, it follows that there exists a time t_0, after which the agents in different components do not interact and exchange their opinions. Therefore, we consider each weakly connected component as a separate and independent instance of the Network-HK model. From Lemma 3.4, the agents in each of these instances reach consensus, thus the Network-HK model converges to a stable state. □

The proof can be generalized in the d-dimensional case, where each agent i holds a d-dimensional opinion $x_i(t) \in [0,1]^d$, with the only change in the proof being that we require $\tau(\boldsymbol{C}) \leq \varepsilon/(2\sqrt{d})$ in (3.3).

While we have proven the convergence of the Network-HK model, no upper or lower bounds on the convergence time are known. The existence of such bounds remains an interesting open question in the field of opinion dynamics.

3.2 The Random-HK Model

A second variation, that departs from the HK model's deterministic nature, is the *Random-HK model* [30]. It is an extension the HK model that, like the Network-HK model, considers the interaction of agents under limited information on the opinions of their neighborhood. However, instead of representing the concept of limited information via an underlying graph, it introduces randomness in the computation of the agents' neighborhoods as a way to limit the spread of information in the network, with each agent learning only a random subset of opinions of the agents in his neighborhood.

3.2.1 Definition

In the Random-HK model, along with the vector of initial opinions $\boldsymbol{x}(0)$ and the agents' confidence ε, we are also given a sampling parameter k. At any time step $t \geq 1$, each agent i samples a random subset K_i of k agents, from all agents, with replacement. Afterwards, agent i computes her neighborhood

$$\mathcal{N}_i(K_i, t, \varepsilon) = \{j : j \in K_i \; and \; |x_i(t-1) - x_j(t-1)| \leq \varepsilon\} \tag{3.4}$$

and updates her opinion to the average of the opinions in $\mathcal{N}_i(K_i, t, \varepsilon)$, as in the HK model. Note that, as with the Network-HK model, (2.26) holds in the Random-HK model as well, with each $\boldsymbol{A}_t$ being an $n \times n$ matrix with the (i,j) element equal to $1/|\mathcal{N}_i(K_i, t, \varepsilon)|$ if $j \in \mathcal{N}_i(K_i, t, \varepsilon)$ and 0 otherwise. Again, we require $i \in \mathcal{N}_i(K_i, t, \varepsilon)$ for every agent i. Also note that $\boldsymbol{A}_t$ corresponds to the adjacency matrix of a directed, unweighted graph, but with normalized rows so that each $\boldsymbol{A}_t$ is stochastic.

3.2.2 Results

The Random-HK model with sampling parameter k converges to a stable state, as is shown by Fotakis et al [30], who introduced the Random-HK model as well. The idea behind their proof is essentially the same as their proof of convergence in the Network-HK model, where some notions are modified to include the inherent randomness of this model. Since the two proofs are quite similar, we will not prove the convergence of the Random-HK model in this thesis, but instead we will define the concepts necessary to the proof in [30] and also state without proof the lemmas that point to the convergence of the Random-HK model. We will begin by defining the notion of an *ε-connected set* of agents that is directly equivalent to the weakly connected set of the Network-HK model.

Definition 3.6 (ε-Connected Set)**.** Let S_1, S_2 be two disjoint sets of agents. Then, $d^t(S_1, S_2) = min_{i \in S_1, j \in S_2} |x_i(t) - x_j(t)|$ denotes their distance at time t. We say that a set of agents $S \subseteq V$ is *ε-connected* at time t, if for any $S' \subset S$ such that $S' \neq \emptyset$, $d^t(S', S \setminus S') \leq \varepsilon$.

Intuitively, ε-connectivity in a set S implies the existence of a path $(q_0 = i, q_1, \ldots, q_{m-1}, q_m = j)$ between any pair of agents $i, j \in S$, such that for each $0 \leq k \leq m-1$, we have $|x_{q_k}(t) - x_{q_{k+1}}(t)| \leq \varepsilon$. We recall our definition of a split between two agents (2.3), and we extend it to define a *break* of a set of agents.

Definition 3.7 (Break)**.** We say that a set of agents S *breaks at time* t, if S is ε-connected at time $t-1$ and it is not ε-connected at time t.

It is important to note that, as with the HK model, once a set of agents S breaks at time t_0 into subsets S' and $S \setminus S'$, the agents in the two subsets no longer communicate and interact with one another, and the subsets behave as independent subinstances of the Random-HK model for all times $t \geq t_0$. This leads us to the first important lemma necessary to prove the convergence of the Random-HK model, that is equivalent to Lemma 3.2 of the Network-HK model.

Lemma 3.8. *Given a set of agents S in the Random-HK model, there is a unique partition of S into ε-connected subsets $S_1, S_2, \ldots, S_m$.*

If we denote by $R_{t_1}^{t_2}$ the event that a break occurs between times t_1 and t_2, and by $\bar{R}_{t_1}^{t_2}$ the event that no breaks occurs between times t_1 and t_2 respectively, we can state the following lemma, equivalent to Lemma 3.4 of the Network-HK model, which shows that in an ε-connected set, all agents converge to a single opinion with probability that tends to 1.

Lemma 3.9. *Let $(k, \varepsilon, \boldsymbol{x}(0))$ be an instance of the Random-HK model. For any $\gamma, \delta > 0$ and any time step $t_1 \geq 0$, there is a time step $t_2 > t_1$ such that $\Pr[\bar{R}_{t_1}^{t_2}] > 0$ and*

$$\Pr\left[\forall i, j \ \ |x_i(t_2) - x_j(t_2)| \leq \gamma \mid \bar{R}_{t_1}^{t_2}\right] \geq 1 - \delta \tag{3.5}$$

Intuitively, the above lemma implies that all ε-connected components will have been formed by time nt_2, and that each ε-connected component reaches consensus separately. Each ε-connected component converges for the same reasons as in the Network-HK model, specifically that all agents influence each other with their opinions, for all times t. Combining Lemmas 3.8 and 3.9, we get that the Random-HK model converges asymptotically to a stable state, with probability that tends to 1.

Theorem 3.10. *The Random-HK model converges to a stable state.*

3.3 The Inertial-HK Model

Another interesting extension of the HK model, which considers the case where agents can move arbitrarily close to the weighted average in their neighborhood at each time step, is the *Inertial-HK model* [31]. This model was introduced to capture the addition of fully-stubborn agents in the HK model. The convergence properties of this model were one of the most significant open questions in the field of opinion dynamics, with extensive simulations pointing to the model's convergence [28, 32]. However, a proof remained elusive for a long time and the efforts of obtaining one led to the innovative energetic approach being utilized for the first time, thus providing us with another powerful mathematical tool for the analysis of coevolutionary models.

3.3.1 Definition

In the Inertial-HK model, instead of being required to move to the mass center of its neighbors at each step, each agent can move towards it by any fraction of length. While all results presented in this section hold for the d-dimensional case, with agent i's opinion being $x_i \in \mathbb{R}^d$, we will focus our analysis on the 1-dimensional case due to its simplicity. We also set $\varepsilon = 1$ in this model, since we can always normalize all agents' opinions by ε, and this simplifies the model's analysis. Therefore, in the Inertial-HK model we have

$$\mathcal{N}_i(t) = \{j : |x_i(t-1) - x_j(t-1)| \leq 1\} \tag{3.6}$$

As in the previous models, we see that $i \in \mathcal{N}_i$. At every time step, each agent i computes her neighborhood and moves to a point that is a convex combination of her previous opinion and the average of the opinions in $\mathcal{N}_i$

$$x_i(t+1) = (1 - \lambda_i(t))\, x_i(t) + \frac{\lambda_i(t)}{|\mathcal{N}_i(t)|} \sum_{j \in \mathcal{N}_i(t)} x_j(t) \tag{3.7}$$

We call $\lambda_i(t) \in [0, 1]$ the *inertia*, which gives the model its name. We see that $\lambda_i(t)$ need not have the same value for all the agents, and is also time-variant. Setting $\lambda_i(t) = 0$ for all times t turns agent i into a fully-stubborn agent. Also note that we can retrieve the original HK model by setting $\lambda_i(t) = 1$ for all agents and all times.

3.3.2 Results

The Inertial-HK model was introduced by Chazelle and Wang [31] in order to prove the convergence of the HK model with fully-stubborn agents, where they utilized the concept of the *kinetic s-energy* of the system, presented in detail in Section 4.5. The kinetic energy of a system was introduced as a generating function for studying averaging processes in dynamic networks [33], and is defined in the d-dimensional case as

$$K(s) = \sum_{t \geq 0} \sum_{i=1}^{n} \|x_i(t+1) - x_i(t)\|_2^s \tag{3.8}$$

We will provide an upper bound on the kinetic s-energy of the Inertial-HK model for $s = 2$ in the 1-dimensional case, and we will utilize this result to prove the asymptotic convergence of the HK model with the addition of fully-stubborn agents.

Intuitively, the bound on the kinetic 2-energy of the Inertial-HK model follows from the fact that $K(2)$ is a quadratic and convex function of $\boldsymbol{x}$ and, as agents move towards each other and form clusters, they tend to move less and $K(2)$ converges asymptotically to its minimum value. If, afterwards, we constrict the possible values of inertia to the binary setting $\lambda_i(t) \in \{0, 1\}$, we get a variation of the HK model with the addition of fully-stubborn agents. The upper bound on $K(2)$ can be used here to proof the existence of a time t_δ, for any arbitrarily small $\delta > 0$, such that for all $t \geq t_\delta$ no agent moves by more than δ. Then, we attempt to show that, after a specific time step, agents are endowed with fixed neighborhoods, thus we essentially have an instance of the DeGroot model (Section 2.1.1). The reason behind this is that, since agents are either non-stubborn or fully-stubborn, they cannot oscillate in and out of neighborhoods forever. Such an alternation for an agent i would imply the existence of another agent j who also oscillates between neighborhoods and causes i's periodic movement. But then, a third

agent has to be the cause of j's periodic movement. Inductively, since our set of agents is finite and the opinions are in 1 dimension, we arrive at a contradiction. In the higher-dimensional case, this last proof is quite different and consists of showing that all agents are either "trapped" by certain fully-stubborn agents and converge asymptotically to a convex combination of their opinions, or become fully-stubborn agents themselves.

Lemma 3.11. *Consider an Inertial-HK system with n agents, whose inertias are uniformly bounded from above by λ_*. Then, the kinetic s-energy of this system satisfies $K(2) \leq \lambda_* n^2/4$.*

Proof. In order to prove the above lemma, we will define a function $C_i(t)$ for each agent i that captures how much each agent moves in the system. Intuitively, we can imagine $C_i(t)$ being the amount of "money" agent i haves at time t. We assign each agent with a certain amount of money at the beginning ($t = 0$), and we introduce a protocol for spending and exchanging it with other agents as time progresses. If we knew ahead of time the total contribution of agent i to the kinetic 2-energy, we could simply set $C_i(0)$ to that amount and let the agent "pay" for her contribution from her own pocket. However, this information is not known beforehand, so we take an initial guess and set up an exchange protocol so that no agent runs out of money. By giving money to their neighbors in a judicious manner, we show how each agent remains in a position to pay for her share of the kinetic 2-energy at each step. The proof is a message-passing protocol that treats $C_i(t)$ as a distributed Lyapunov function, and simulates its update. In the beginning, we assign agent i

$$C_i(0) = \sum_{j=1}^{n} \min\Big\{ |x_i(0) - x_j(0)|^2,\, 1 \Big\} \tag{3.9}$$

units of money. Intuitively, agent i will have to spend an amount of money each turn that represents how much he moved at that time step. Furthermore, agent i gives money to agent j as a way to pay for the cost that j incurs from the fact that i moved from his previous position. Since j also moves, concurrently with i, the inverse exchange from j to i happens as well. In addition, because agents move towards each other and form clusters, once i exits the neighborhood of some j, he stops exchanging money with that agent. Therefore, as clusters are formed, agents move less, interact less and the flow of money decreases. That is the idea behind this approach to proving that $K(2)$ is bounded.

Before we proceed, we will define some useful quantities that will simplify the equations below. For any two agents i and j we have

i. $\Delta_i = x_i(t+1) - x_i(t)$

ii. $d_{ij} = x_i(t) - x_j(t)$

iii. $d'_{ij} = x_i(t+1) - x_j(t+1)$

Our message-passing protocol consists of two rules, applied to every agent i at any time step $t \geq 0$

- Agent i spends $(\Delta_i + \Delta_j)^2$ units of money for every $j \in \mathcal{N}_i(t)$, at each time t, and gives to agent j an amount equal to $2(d_{ij} - \Delta_j)\Delta_j$. Note that, since $i \in \mathcal{N}_i$, agent i spends at least $4{\Delta_i}^2$ units of money at each time t.
- For every agent j that becomes, or ceases to be, a neighbor of i at time $t+1$, agent i spends $|{d'_{ij}}^2 - 1|$.

Before we continue, we will first simply the notation. For the remainder of this proof, we will denote by $\mathcal{N}_i$ agent i's neighborhood at time t, since we will focus at two specific times t and $t+1$. For the same reasons, we will denote $\lambda_i(t)$ by λ_i. We will also make a distinction in the neighborhood of agent i, denoting by $\mathcal{N}_i^{in}$ the set of agents that are neighbors of i at time $t+1$, but not at time t, and by $\mathcal{N}_i^{out}$ the set of agents that are neighbors of i at time t, but not at time $t+1$. Focusing on a single agent, we analyze the cash flow at time t

$$\begin{aligned} C_i(t+1) - C_i(t) = &- \sum_{j \in \mathcal{N}_i} (\Delta_i + \Delta_j)^2 + 2 \sum_{j \in \mathcal{N}_i} (d_{ji} - \Delta_i)\Delta_i \\ &- 2 \sum_{j \in \mathcal{N}_i} (d_{ij} - \Delta_j)\Delta_j - \sum_{j \in \mathcal{N}_i^{in} \cup \mathcal{N}_i^{out}} |{d'_{ij}}^2 - 1| \end{aligned} \tag{3.10}$$

We also have that

$$(d_{ji} - \Delta_i)\Delta_i - (d_{ij} - \Delta_j)\Delta_j = d_{ji}\Delta_i - \Delta_i^2 - d_{ij}\Delta_j + \Delta_j^2$$

which, since $d_{ji} = -d_{ij}$, is equal to

$$- d_{ij}(\Delta_i + \Delta_j) + \Delta_j^2 - \Delta_i^2 = d_{ij}(\Delta_i - \Delta_j) - 2d_{ij}\Delta_i + \Delta_j^2 - \Delta_i^2 \tag{3.11}$$

Combining (3.10) and (3.11), we get

$$\begin{aligned} C_i(t+1) - C_i(t) = & \sum_{j \in \mathcal{N}_i} \Big\{ 2d_{ij}(\Delta_i - \Delta_j) - 4d_{ij}\Delta_i - (\Delta_i^2 + 2\Delta_i\Delta_j + \Delta_j^2) \\ & +2\Delta_j^2 - 2\Delta_i^2 \Big\} - \sum_{j \in \mathcal{N}_i^{in} \cup \mathcal{N}_i^{out}} |{d'_{ij}}^2 - 1| \\ = & \sum_{j \in \mathcal{N}_i} \Big\{ 2d_{ij}(\Delta_i - \Delta_j) - 4d_{ij}\Delta_i + (\Delta_i^2 - 2\Delta_i\Delta_j + \Delta_j^2) \\ & -4\Delta_i^2 \Big\} - \sum_{j \in \mathcal{N}_i^{in} \cup \mathcal{N}_i^{out}} |{d'_{ij}}^2 - 1| \end{aligned} \tag{3.12}$$

Furthermore, from (3.7) we have

$$\Delta_i = \lambda_i \sum_{j \in \mathcal{N}_i} \frac{x_j(t)}{|\mathcal{N}_i|} - \lambda_i x_i(t)$$

which gives us

$$\sum_{j \in \mathcal{N}_i} d_{ij} = -\lambda_i^{-1} |\mathcal{N}_i| \Delta_i \tag{3.13}$$

Combining now (3.12) and (3.13), we get

$$\begin{aligned} C_i(t+1) - C_i(t) = & \sum_{j \in \mathcal{N}_i} \Big\{ 2d_{ij}(\Delta_i - \Delta_j) - 4d_{ij}\Delta_i + (\Delta_i - \Delta_j)^2 \Big\} - 4|\mathcal{N}_i|\Delta_i^2 \\ & - \sum_{j \in \mathcal{N}_i^{in} \cup \mathcal{N}_i^{out}} |{d'_{ij}}^2 - 1| \\ = & \sum_{j \in \mathcal{N}_i} \Big\{ 2d_{ij}(\Delta_i - \Delta_j) + (\Delta_i - \Delta_j)^2 \Big\} - 4(\lambda_i^{-1} - 1)|\mathcal{N}_i|\Delta_i^2 \\ & - \sum_{j \in \mathcal{N}_i^{in} \cup \mathcal{N}_i^{out}} |{d'_{ij}}^2 - 1| \end{aligned} \tag{3.14}$$

Dividing by λ_i in (3.13) is possible, due to $\lambda_i = 0$ implying $\Delta_i = 0$. Therefore, in this case, i is a fully-stubborn agent and $C_i(t+1) - C_i(t) = 0$ for all times t which means that agent i never runs out of money. Since $d'_{ij} - d_{ij} = \Delta_i - \Delta_j$, we have

$$2d_{ij}(\Delta_i - \Delta_j) + (\Delta_i - \Delta_j)^2 = 2d_{ij}d'_{ij} - 2d_{ij}^2 + (d'_{ij} - d_{ij})^2 = d'^2_{ij} - d_{ij}^2 \tag{3.15}$$

Thus, from (3.14) and (3.15), we get

$$C_i(t+1) - C_i(t) = \sum_{j \in \mathcal{N}_i} \left\{ d_{ij}'^2 - d_{ij}^2 \right\} - \sum_{j \in \mathcal{N}_i^{in} \cup \mathcal{N}_i^{out}} |{d'_{ij}}^2 - 1| - 4(\lambda_i^{-1} - 1)|\mathcal{N}_i|\Delta_i^2 \tag{3.16}$$

The first two sums in the equation above can be combined, if we note that

- For all $j \in \mathcal{N}_i^{out}$, we have $d_{ij} \leq 1$ and $d'_{ij} > 1$. Therefore $d_{ij}'^2 - d_{ij}^2 - |{d'_{ij}}^2 - 1| = 1 - d_{ij}^2$.
- For all $j \in \mathcal{N}_i^{in}$, we have $d_{ij} > 1$ and $d'_{ij} \leq 1$. Therefore these agents only participate in the seccond summand, with their contribution being $-|d_{ij}'^2 - 1| = d_{ij}'^2 - 1$.
- For all other agents $j \in \mathcal{N}_i(t) \cap \mathcal{N}_i(t+1)$, we have $d_{ij} \leq 1$ and $d'_{ij} \leq 1$. Therefore, these agents only participate in the first summand, with their contribution being $d_{ij}'^2 - d_{ij}^2$.

Thus, (3.16) can be written as

$$C_i(t+1) - C_i(t) = \sum_{j=1}^{n} \min\{d_{ij}'^2, 1\} - \sum_{j=1}^{n} \min\{d_{ij}^2, 1\} - 4(\lambda_i^{-1} - 1)|\mathcal{N}_i|\Delta_i^2 \tag{3.17}$$

Iterating the above equation, and using the fact that $\lambda_i \leq \lambda_*$, it follows that

$$C_i(t) \geq \sum_{j=1}^{n} \min\{d_{ij}^2, 1\} + 4\big(\lambda_*^{-1} - 1\big) \sum_{k=0}^{t-1} (x_i(k+1) - x_i(k))^2 \tag{3.18}$$

Being its own neighbor, agent i spends at least $4\Delta_i^2$ money at each step. Summing up over all the agents, this amounts to $4K(2)$. This shows that the initial injection of money allows the system to spend $4K(2)$ and still be left with $4\big(\lambda_*^{-1} - 1\big)K(2)$. The proof of the lemma now follows directly from the fact that the initial injection of money is at most n^2. □

Note that, at each time step t, the set of neighbors $\mathcal{N}_i(t)$ forms a directed graph G_t, called the *communication network*, that changes with time. The bound on the kinetic 2-energy shows that the model slows down to a crawl but it is not enough to prove convergence, as an agent moving along a circle by $1/t$ at time t contributes finitely to the kinetic 2-energy of the system yet travels an infinite distance. However, we can

utilize this bound on the kinetic 2-energy to prove that the HK model with fully-stubborn agents, where each agent's inertia is either 0 or 1, always converges asymptotically. In the 1-dimensional case, which we will present here, the communication network settles on a fixed graph for all initial conditions. In the d-dimensional case with $d > 1$, the communication network converges for all initial conditions outside a set of measure zero, which implies that a pertubation of the fully-stubborn agents by an arbitrarily small amount in the beginning ensures that the system will converge to a fixed configuration and the communication network will settle on a fixed graph.

Theorem 3.12. *The Inertial-HK model, where each agent's inertia is either* 0 *or* 1*, converges asymptotically, in the case of* $d = 1$*, to a fixed-point configuration and the communication graph settles on a fixed graph for all initial conditions.*

Proof. The upper bound on the kinetic 2-energy that we get by Lemma 3.11 shows that, for any arbitrarily small $\varepsilon > 0$, there exists a time step t_ε such that no agent moves by a distance of more than ε at any time $t \geq t_\varepsilon$. Consider a fixed time $t_0 > t_\varepsilon$, and let, for brevity, $x_i = x_i(t_0)$ and $\mathcal{N}_i = \mathcal{N}_i(t_0)$ for each agent i. We use primes and double primes to indicate the equivalent quantities for times $t_0 + 1$ and $t_0 + 2$.

We will introduce notation for four different sets of agents

- Let L_i^{in} be the set of agents located at $x_i - 1 - \mathcal{O}(\varepsilon)$ at time t_0 and at $x_i - 1 + \mathcal{O}(\varepsilon)$ at time $t_0 + 1$. This set consists of the agents that joined i's neighborhood at $t_0 + 1$ from the left side of agent i.
- Let L_i^{out} be the set of agents located at $x_i - 1 + \mathcal{O}(\varepsilon)$ at time t_0 and at $x_i - 1 - \mathcal{O}(\varepsilon)$ at time $t_0 + 1$. This set consists of the agents that left i's neighborhood at $t_0 + 1$ from the left side of agent i.
- Let R_i^{in} be the set of agents located at $x_i + 1 + \mathcal{O}(\varepsilon)$ at time t_0 and at $x_i + 1 - \mathcal{O}(\varepsilon)$ at time $t_0 + 1$. This set consists of the agents that joined i's neighborhood at $t_0 + 1$ from the right side of agent i.
- Let R_i^{out} be the set of agents located at $x_i + 1 - \mathcal{O}(\varepsilon)$ at time t_0 and at $x_i + 1 + \mathcal{O}(\varepsilon)$ at time $t_0 + 1$. This set consists of the agents that left i's neighborhood at $t_0 + 1$ from the right side of agent i.

It is obvious that all the sets introduced above are disjoint, and their union is the symmetric difference between $\mathcal{N}_i$ and $\mathcal{N}_i'$. The locations x_i' and x_i'' of agent i at times $t_0 + 1$ and $t_0 + 2$ are given by

$$|\mathcal{N}_i|x_i' = \sum_{j\in\mathcal{N}_i\cap\mathcal{N}_i'} x_j + \sum_{j\in L_i^{out}\cup R_i^{out}} x_j$$
$$|\mathcal{N}_i'|x_i'' = \sum_{j\in\mathcal{N}_i\cap\mathcal{N}_i'} x_j' + \sum_{j\in L_i^{in}\cup R_i^{in}} x_j'$$

Since all x_k' and x_k'' are of the form $x_k \pm \mathcal{O}(\varepsilon)$, subtracting the two identities above shows that

$$(|\mathcal{N}_i'| - |\mathcal{N}_i|)x_i = (|L_i^{in}| - |L_i^{out}|)(x_i - 1) + (|R_i^{in}| - |R_i^{out}|)(x_i + 1) \pm \mathcal{O}(\varepsilon n) \tag{3.19}$$

The dynamics is translation-invariant, thus we can set $x_i = 0$. If we choose a small enough ε, the integrality of the set cardinalities implies that the net flow of neighbors on the left of agent i is the same as it is on the right

$$|L_i^{out}| - |L_i^{in}| = |R_i^{out}| - |R_i^{in}| \tag{3.20}$$

Let us now focus on the agents that are undergoing a change of neighbors between times t_0 and t_0+1. Among these agents, we choose the one that ends up the furthest to the right at time t_0+1, breaking ties by picking the agent with the largest index. We call this agent i, and distinguish between two cases

i. $x_i' \geq x_i$: Agent i moves to the right. Thus, no agent of R_i^{out} can be fully-stubborn, since the fully-stubborn agents do not move and cannot leave the neighborhood of an agent moving to the right from his right side. Also, no agent of R_i^{out} can be mobile, since, with the ordering of agents being preserved in the HK model (Section 2.2.1), this would imply the existence of an agent that undergoes a change of neighbors at time t_0+1 and lands to the right of i at time t_0+1, in contradiction with the definition of i. Therefore, R_i^{out} must be empty. This in turn implies that L_i^{in} is not empty, since i undergoes a change of neighbors at time t_0+1, and this means that not all four terms in (3.20) can be zero. Since agent i is not moving left, neither is any agent j of L_i^{in}. Its set $\mathcal{N}_j$ of neighbors changes between times t_0 and t_0+1 and R_j^{out} is empty. The latter is true, since $\mathcal{N}_j$ cannot lose any fully-stubborn agents to the right and, in addition, any mobile agent k in R_j^{out} is to the left of i at time t_0, stays to the left of i at time t_0+1 by conservation of ranks, therefore, since i is in $\mathcal{N}_j$, k must also be in $\mathcal{N}_j$. The argument so far uses the rightmost status of agent i only to assert that R_i^{out} is empty. This means that

we can now forget about agent i, choose an agent in L_i^{in}, and proceed inductively, eventually reaching a contradiction.

ii. $x_i' < x_i$: Agent i moves to the left. However, note here that our previous argument never uses time directionality, so we can exchange the role of t_0 and $t_0 + 1$, which implies that now $x_i' > x_i$. We must also swap the superscripts *in* and *out* and instead of choosing i as the agent landing furthest to the right, by symmetry we choose agent i as the agent starting the furthest to the right. Since in the HK model the orderings of the agents are being preserved, this does not violate our previous argument and we can use it to reach a contradiction in this case as well.

We conclude that, at time t_0, all four terms of (3.20) are equal to zero, for all agents. Thus, each agent now has a fixed set of neighbors, so the dynamics is specified by the powers of a fixed stochastic matrix with positive diagonal. Therefore, the communication network is a fixed graph, and we have an instance of the DeGroot model, which is well known to converge, as we demonstrated in Section 2.1.1. The system is attracted to a fixed point at an exponential rate, but we can provide no bound on the time it takes to fall into that basin of attraction. □

While we have shown that the HK model with fully-stubborn agents always converges asymptotically to a stable state, no effective upper bound on the convergence time is known, and such an upper bound remains an interesting open question.

3.4 The Asymmetric k-NN Model

We will now consider a variant of the HK model that differs significantly from all other models, the *Asymmetric k-Nearest Neighbor (k-NN) model* [34]. In the k-NN model, agent i forms directed links to his k nearest neighbors, thus each agent's neighborhood consists of exactly k agents. In this model, agent i's cost function is not continuous in $\boldsymbol{x}_{-i}$, which denotes the vector of all agents' opinions except for agent i, therefore the k-NN model need not admit to a pure Nash equilibrium. Indeed, we will show that this model need not converge, even for the simple case of $k = 1$.

3.4.1 Definition

In the Asymmetric k-NN model, each agent i holds a permanent, intrinsic opinion s_i, and expresses an opinion x_i that could be different from s_i, as with the FJ model in Chapter 2. At any time t, each agent i computes her neighborhood $\mathcal{N}_i(\boldsymbol{x})$, forming directed links to the k agents with smallest distance $|x_j(t) - s_i|$ and breaking ties in a

consistent fashion. Furthermore, all agents assign a weight ρ to their intrinsic opinion when they update their expressed opinions. In this model, ρ is constant in time and uniform for all agents. Each agent i incurs a cost at time t equal to

$$C_i(x_i, \boldsymbol{x}_{-i}) = \sum_{j \in \mathcal{N}_i(\boldsymbol{x})} (x_i(t) - x_j(t))^2 + \rho k (x_i(t) - s_i)^2 \tag{3.21}$$

In contrast with all previous models, $i \notin \mathcal{N}_i(\boldsymbol{x})$ in the k-NN model, therefore i's neighborhood depends only on $\boldsymbol{x}_{-i}$. In order to minimize her cost at time t, agent i sets x_i to be the weighted average of the opinions in $\mathcal{N}_i(\boldsymbol{x})$ and s_i

$$x_i(t+1) = \frac{\sum_{j \in \mathcal{N}_i(\boldsymbol{x})} x_j(t) + \rho k s_i}{k(\rho + 1)} \tag{3.22}$$

3.4.2 Results

In this section, we will show that the Asymmetric k-NN model need not converge to a pure strategy Nash equilibrium, even for $k = 1$, with a simple counterexample. This was proven by Bhawalkar et al [34], who also introduced this model. Before we continue, we will simplify our notation for $k = 1$. Let $l_i(t) = argmin_{j \neq i} |x_j(t) - s_i|$ denote the nearest neighbor of i at time t, with ties breaking in a consistent fashion. The, the cost agent i incurs at time t is

$$C_i(x_i, \boldsymbol{x}_{-i}) = (x_i(t) - x_{l_i}(t))^2 + \rho (x_i(t) - s_i)^2 \tag{3.23}$$

Agent i minimizes her cost at time t if she sets x_i equal to

$$x_i(t+1) = \frac{x_{l_i}(t) + \rho s_i}{\rho + 1} \tag{3.24}$$

We can now prove the following theorem

Theorem 3.13. *Consider an instance of the 1-NN model, where $n = 3$, $\rho = 1$ and the agents' intrinsic opinions are $s_1 = 0$, $s_2 = 1/2$ and $s_3 = 1$. This game does not admit a pure strategy Nash equilibrium.*

Proof. First of all, note that $\boldsymbol{x} \in [0, 1]^3$, since there is no point for any agent to express an opinion outside the minimum and maximum values of $\boldsymbol{s}$. Suppose that a pure Nash equilibrium exists, with $\boldsymbol{x} = [a, b, c]^T$ in this equilibrium. Firstly, note that c cannot be less than both a and b, since c is either $(1 + a)/2 \geq a$ or $(1 + b)/2 \geq b$. Similarly, a cannot be greater than both b and c, since a is either $b/2 \leq b$ or $c/2 \leq c$. These imply

that $a \leq c$ and, in particular, $a < c$, since if $a = c$, at least one of them has a feasible deviation. In addition, if $b \leq a$, then the first agent points to the second agent and has a feasible deviation, and if $b \geq c$, the third agent points to the second agent and has a feasible deviation. Therefore, the ordering between the agents' expressed opinions must be $a < b < c$.

Since the ordering is $a < b < c$, the first agent points to the second agent and the third agent also points to the second agent. Therefore, $a = b/2$ and $c = (1 + b)/2$. We distinguish between two cases, one where the second agent points to the first agent, and one where the second agent points to the third agent.

- If the second agent points to the first agent, we have that $b = (1/2 + a)/2 = (1 + b)/4$. Solving this for b, we get $b = 1/3$, which, in turn, gives $a = 1/6$ and $c = 2/3$. Since c is closer to $1/2$ than a, the second agent should instead point to the third agent.
- If the second agent points to the third agent, we have that $b = (1/2 + c)/2 = (2 + b)/4$. Solving this for b, we get $b = 2/3$, which, in turn, gives $a = 1/3$ and $c = 5/6$. Since a is closer to $1/2$ than c, the second agent should instead point to the first agent.

Thus, a pure Nash equilibrium does not exist in this game. □

3.5 The Generalized Asymmetric Model

In this section we will consider a generalization of the HK model. Specifically, we will define a game where each weight assigned by agent i to agent j depends on the distance of all agents' expressed opinions from i's intrinsic opinion. Moreover, j's influence over i is asymmetric, and minimal assumptions are being made on the agents' weights. However, in contrast with the previous model, the agents' cost functions are continuous in this model and Rosen's theorem can be applied to guarantee the existence of a pure Nash equilibrium. This model is called the *Generalized Asymmetric model* [34], of which the k-NN model is a special case.

3.5.1 Definition

In the Generalized Asymmetric model, as with the previous model, each agent i holds a permanent, intrinsic opinion s_i, and expresses an opinion x_i that could be different from s_i. However, in this model, the vector of expressed opinions $\boldsymbol{x}_{-i}$ along with s_i determines the strength of i's friendship with all other agents. We define the distance of agent j's

expressed opinion from agent i's intrinsic opinion at time t as $d_j^i(t) = |x_j(t) - s_i|$, for $j \neq i$. The weight that i assigns to j's expressed opinion is $q_{ij}(\boldsymbol{x}_{-i}, s_i) = F_i(d_j^i(t), d_{-i,-j}^i(t))$, where F_i is a continuous function that decreases as $d_j^i(t)$ increases and increases as $d_{-i,-j}^i(t)$ increases. Intuitively, if we freeze all other agents besides j, i assigns more weight to j if their distance decreases, and less if it increases. Additionally, if we freeze agents i and j, i assigns more weight to j if the other agents move away from s_i and less if they move towards s_i. Also, each agent has a different weight ρ that she assigns to her own intrinsic opinion when she updates her expressed opinion.

The cost that agent i incurs in the Generalized Asymmetric model is

$$C_i(x_i, \boldsymbol{x}_{-i}) = \sum_{j \neq i} q_{ij}(t)(x_i(t) - x_j(t))^2 + \rho_i(x_i(t) - s_i)^2 \tag{3.25}$$

We see that each agent has a unique dominant strategy at any time step, to express the opinion that minimizes her cost. Therefore, each agent has a best response

$$x_i(t+1) = \frac{\sum_{j \neq i} q_{ij}(t) x_j(t) + \rho_i s_i}{\sum_{j \neq i} q_{ij}(t) + \rho_i} \tag{3.26}$$

that minimizes her cost, if all other agents do not move at time $t + 1$.

3.5.2 Results

Unfortunately, not much is known about the convergence properties of the Generalized Asymmetric model. For continuous cost functions C_i, we have the next theorem which follows from Rosen's theorem.

Theorem 3.14. *The Generalized Asymmetric model admits to a pure strategy Nash equilibrium when the cost functions C_i are continuous.*

Proof. Since q_{ij} is independent of x_i and only depends on $\boldsymbol{x}_{-i}$ and s_i, the function $q_{ij}(\boldsymbol{x}_{-i}, s_i)(x_i(t) - x_j(t))^2$ is convex in x_i. Therefore, $C_i(x_i, \boldsymbol{x}_{-i})$ is convex in x_i and also continuous in $\boldsymbol{x}$, by our assumption. This implies that the Generalized Asymmetric model is a concave game (Section 4.3.1) and, from Rosen's theorem, admits to a pure strategy Nash equilibrium [35]. □

While we have proved that there exists a Nash equilibrium in the Generalized Asymmetric model, it is not known whether the model always converges to the Nash equilibrium. The convergence of the Generalized Asymmetric model remains one of the most interesting open questions on non-linear models, and in the field of opinion dynamics in general.

Chapter 4

Model Analysis Toolbox

The purpose of this chapter is to present several mathematical tools used in the analysis of the different models in the field of opinion dynamics. We will present the ideas behind these theorems and specify their overall contribution in the field. While there are many interesting scientific results that utilize ideas not presented in this chapter, we believe the collection of tools presented here consists of some of the most fundamental and important ideas that were used to provide us with the most significant results. This chapter is quite important as, in our opinion, one needs to be familiarized with almost all the following concepts in order to undertake research in the field of opinion dynamics.

We begin by defining the potential games as those that admit a specific potential function and show that the point where this function reaches its optimum value is exactly the Nash equilibrium of a system. We continue by presenting certain fixed-point theorems that are utilized to prove the existence of a Nash equilibrium, and maybe offer some ideas on how to prove convergence. Then, we focus our attention on concave games, where the cost functions of the agents are concave, and state one of the most important theorems about them, proved by Rosen [35], before continuing with gradient descent-like methods that are useful for minimizing specific functions. Finally, we present the innovative ideas of the energy approach to a system, that can be used to circumvent several problems arising with step-by-step methods.

4.1 Potential Functions

The concept of a potential function was first used in the analysis of congestion games. Specifically, Rosenthal proved, in 1973, that every congestion game has a function which can be used to prove the existence of, and sometimes convergence to, a Nash equilibrium [36]. We call this function the *potential function* of the game and the idea behind the concept is to provide a sense of quantifiable distance from the game's equilibrium point.

In addition, games that possess such functions are call *potential games*. Monderer and Shapley utilized potential functions in 1996 to prove the converse; for every potential game, there exists a congestion game with the same potential function [37].

While the concept of the potential function was, at first, closely associated with congestion games, it has since grown as a mathematical tool and is utilized in general optimization problems, due to its simplicity and usefulness. In our field, we can consider the agents, each one trying selfishly to minimize her own cost, as having a global, unified potential function that they are working together to optimize. Thus, the potential function in opinion dynamics is closely associated with the cost that each agent incurs in the model.

We will continue by defining the basic types of potential functions before presenting one that has provided interesting results in the field. Consider a model with n agents that can express any opinion $x_i \in \mathbb{R}$, and have cost functions $C_i : \mathbb{R}^n \to \mathbb{R}$.

Definition 4.1 (Exact Potential Function)**.** If, in our model, there exists a function $\Phi : \mathbb{R}^n \to \mathbb{R}$ such that for every agent i and every x_i, x_i'

$$\Phi(x_i', \boldsymbol{x}_{-i}) - \Phi(x_i, \boldsymbol{x}_{-i}) = C_i(x_i', \boldsymbol{x}_{-i}) - C_i(x_i, \boldsymbol{x}_{-i}) \tag{4.1}$$

then Φ is called an *exact potential function* of our model.

Intuitively, the exact potential function has the property that when any agent i switches from an expressed opinion x_i to x_i' with all other agents fixed, the change in the potential function is exactly equal to the change in i's cost. We can generalize the concept by introducing a weight that makes the two values not exactly equal but proportional.

Definition 4.2 (Weighted Potential Function)**.** If, in our model, there exists a function $\Phi : \mathbb{R}^n \to \mathbb{R}$ and a vector $\boldsymbol{w} \in \mathbb{R}^n_{++}$, such that for every agent i and every x_i, x_i'

$$\Phi(x_i', \boldsymbol{x}_{-i}) - \Phi(x_i, \boldsymbol{x}_{-i}) = w_i(C_i(x_i', \boldsymbol{x}_{-i}) - C_i(x_i, \boldsymbol{x}_{-i})) \tag{4.2}$$

then Φ is called a *weighted potential function* of our model.

We can further generalize the concept by merely requiring the signs of the differences in the values above to be equal.

Definition 4.3 (Ordinal Potential Function)**.** If, in our model, there exists a function $\Phi : \mathbb{R}^n \to \mathbb{R}$, such that for every agent i and every x_i, x_i'

$$C_i(x_i', \boldsymbol{x}_{-i}) - C_i(x_i, \boldsymbol{x}_{-i}) > 0 \iff \Phi(x_i', \boldsymbol{x}_{-i}) - \Phi(x_i, \boldsymbol{x}_{-i}) > 0 \tag{4.3}$$

then Φ is called an *ordinal potential function* of our model.

Intuitively, in a game with an ordinal potential function with all other agents fixed, when agent i's cost increases, Φ increases as well, and when i's cost decreases, Φ decreases as well, for any agent i. Before we continue, we will present a significant result on potential functions concerning their relation to pure Nash equilibria.

Theorem 4.4. *Every potential game admits to at least one pure strategy Nash equilibrium. Furthermore, if Φ is the game's potential function, every Nash equilibrium is a local optimum of Φ.*

This theorem was proved by Monderer and Shapley in 1996 [37]. However, we will skip the proof as it is fairly simple and follows directly from our definition of the potential function and the finiteness of each agent's sequence of improvement steps. It is understood now that potential functions are extremely helpful tools in optimization problems, when they exist, and, besides guaranteeing the existence of a pure Nash equlibrium in our model, they also provide us with significant insight and interesting convergence properties.

4.1.1 Application to Opinion Dynamics

In this section, we continue by presenting the link between potential games and opinion dynamics. The well-studied properties of potential functions have been utilized to provide deep insight and results in the field [9, 11]. It should be clear by now that potential functions are extremely useful for the analysis of an opinion formation model, when they exist. However, the important question remains open; when do opinion formation models admit to a potential function? While this question is not resolved for the general case of ordinal potential functions, we present here a necessary condition for an opinion formation model to admit to an exact potential function.

Theorem 4.5. *Consider an opinion formation model where each agent i expresses opinion $x_i \in \mathbb{R}$ and has a cost function $C_i : \mathbb{R}^n \to \mathbb{R}$ that is continuous and twice differentiable. Then, the model admits to an exact potential function if and only if for any two agents i and j*

$$\frac{\partial^2 C_i(\boldsymbol{x})}{\partial x_i \partial x_j} = \frac{\partial^2 C_j(\boldsymbol{x})}{\partial x_i \partial x_j} \tag{4.4}$$

We continue by proving that the undirected FJ model (Section 2.1.2) admits a potential function used to provide tight bounds on the Price of Anarchy [11]. Indeed, the cost functions in the undirected FJ model satisfy the necessary condition of Theorem 4.5, since $\frac{\partial^2 C_i(\boldsymbol{x})}{\partial x_i \partial x_j} = -2w_{ij}$, $\frac{\partial^2 C_j(\boldsymbol{x})}{\partial x_i \partial x_j} = -2w_{ji}$ and $w_{ij} = w_{ji}$. Therefore, the undirected FJ model admits to an exact potential function as stated by the theorem below.

Theorem 4.6. *Consider an instance of the undirected FJ model with n agents that have intrinsic opinions $\boldsymbol{s}$ and expressed opinions $\boldsymbol{x}$. Then, this model admits to an exact potential function*

$$\Phi(\boldsymbol{x}) = \sum_{\{i,j\} \in E(G)} w_{ij}(x_i - x_j)^2 + \sum_{i=1}^{n} w_{ii}(x_i - s_i)^2 \tag{4.5}$$

Proof. Let i be an agent that deviates from his expressed opinion x_i to x_i', while all other agents remain fixed. The cost that i incurs is $C_i(x_i, \boldsymbol{x}_{-i}) = \sum_{j \in \mathcal{N}_i} w_{ij}(x_i - x_j)^2 + w_{ii}(x_i - s_i)^2$. Therefore, the difference in i's cost from his deviation is

$$\begin{aligned} C_i(x_i', \boldsymbol{x}_{-i}) - C_i(x_i, \boldsymbol{x}_{-i}) = &\sum_{j \in \mathcal{N}_i} w_{ij}\big((x_i' - x_j)^2 - (x_i - x_j)^2\big) \\ &+ w_{ii}\big((x_i' - s_i)^2 - (x_i - s_i)^2\big) \end{aligned}$$

If we look at the difference in value of the potential function before and after the deviation, we have that all summands of the form $w_{uv}(x_u - x_v)^2$ along with those of the form $w_{uu}(x_u - s_u)^2$, where $u, v \neq i$, are negated since both u and v remain fixed. Therefore, we get

$$\begin{aligned} \Phi(x_i', \boldsymbol{x}_{-i}) - \Phi(x_i, \boldsymbol{x}_{-i}) = &\sum_{\{i,j\} \in E(G)} w_{ij}\big((x_i' - x_j)^2 - (x_i - x_j)^2\big) \\ &+ w_{ii}\big((x_i' - s_i)^2 - (x_i - s_i)^2\big) \end{aligned}$$

Since we have an instance of the undirected FJ model, $w_{ij} = w_{ji}$ for any two agents i, j. This implies that the sets $A = \big\{j : \{i,j\} \in E(G)\big\}$ and $\mathcal{N}_i$ are equal, hence the two differences above are equal, and Φ is an exact potential function of the FJ model. □

4.2 Fixed-Point Theorems

In this section, we analyze one of the most fundamental mathematical concepts used to prove the existence, and sometimes uniqueness, of a Nash equilibrium in game theory,

the *fixed point*. We present two of the most significant fixed-point theorems, by Brouwer and Kakutani, and show their relation to game theory and opinion dynamics in particular. While fixed-point theorems appear in many different regions of mathematics and their usefulness cannot be overstated, they hold a distinguished place in the field of game theory, as they were used by Nash in his development of the Nash equilibrium as a solution concept for non-cooperative games. We begin by defining the concept of a fixed point in the most general setting.

Definition 4.7 (Fixed Point)**.** Consider a function $F : \mathbb{R}^n \to \mathbb{R}^n$. If there exists a point $\boldsymbol{x_0} \in \mathbb{R}^n$ such that $F(\boldsymbol{x_0}) = \boldsymbol{x_0}$, then $\boldsymbol{x_0}$ is called a *fixed point* of F.

In general, a fixed-point theorem is a result stating that a function F will have at least one fixed point under certain conditions on F that can be stated in general terms. Although there exist a significant number of fixed-point theorems in mathematics, only a handful are of interest in game theory. We present two that we consider the most significant, which played a central role in the proof of existence of general equilibrium in market economies by Arrow and Debreu [38] and in the proof of existence of a mixed Nash equilibrium in every finite game for any number of players by Rosen [35], starting with Brouwer's fixed-point theorem.

4.2.1 Brouwer's Fixed-Point Theorem

Brouwer's fixed-point theorem [39] stands out among hundreds of others due to its broad range of applications across numerous fields of mathematics. In its original field, this result is one of the key theorems characterizing the topology of Euclidean spaces, which gives it a place among the fundamental theorems of topology. Here, we present a simple version of the theorem in the plane and subsequently generalize it to any convex compact set.

Theorem 4.8. *Let $\mathcal{D} = \{(x, y) \in \mathbb{R}^2 : (x - a)^2 + (y - b)^2 \leq r\}$ be a closed disk in $\mathbb{R}^2$, with center (a, b) and radius r, and $f : \mathcal{D} \to \mathcal{D}$ a continuous function. Then, f has at least one fixed point.*

While the theorem's proof is somewhat complicated, it is surprisingly easy to prove in one dimension, thus we will state the proof for a continuous function f defined on a closed interval $[a, b] \subset \mathbb{R}$ that takes values on the same interval.

Proof. Consider the function $g(x) = f(x) - x$. We have that $g(a) \geq 0$ and $g(b) \leq 0$. Then, by the intermediate value theorem, there exists a point $x_0 \in [a, b]$ such that $g(x_0) = 0$. Therefore, $f(x_0) = x_0$, and x_0 is a fixed point of f. □

Intuitively, Theorem 4.8 implies that if one stirs a cup of coffee to dissolve a lump of sugar, there is always a point without motion. However, this example is not a perfect one as it does not demonstrate the non-uniqueness of the fixed point. A better example is if one takes two identical horizontal sheets, crumple and flatten one of them and then place it on top of the other. Brouwer's fixed-point theorem then implies that there exists a point on the crumpled sheet that is in the same place as on the other sheet.

We continue by generalizing Theorem 4.8 to any convex compact set in the Euclidean space.

Theorem 4.9. *Let K be a convex compact (closed and bounded) subset of a Euclidean space, and $f : K \to K$ a continuous function. Then, f has at least one fixed point.*

We should note that each of the preconditions necessary by the theorem is very important, since the violation of any of them renders the theorem unprovable. Indeed, we provide a counterexample for every case

- K is convex and closed, but not bounded:

  Consider the function $f(x) = x + 1$ from $\mathbb{R}$ to itself. Since $\mathbb{R}$ is convex and closed but not bounded, the theorem does not hold. Indeed, as f shifts each point to the right, it cannot have a fixed point.

- K is convex and bounded, but not closed:

  Consider the function $f(x) = \frac{x+1}{2}$ from the open interval $(-1, 1)$ to itself. Since $(-1, 1)$ is convex and bounded, but not closed, the theorem does not hold. Indeed, as f again shifts each point to the right, it cannot have a fixed point. Note that f has a fixed point in the closed interval $[-1, 1]$, namely $f(1) = 1$.

- K is compact (closed and bounded), but not convex:

  Consider the function $f(r, \theta) = (r, \theta + \pi/4)$ in polar coordinates, from the unit circle to itself. Since the unit circle is compact but not convex (as it has a hole), the theorem does not hold. Indeed, as f shifts each point by 45 degrees in the circle, it cannot have a fixed point. Note that f has a fixed point in the unit disk, namely the origin $(0, 0)$.

In addition, it should be noted that Brouwer's fixed-point theorem and Sperner's lemma [40], an important result in combinatorics and very useful in game theory, are equivalent, as assuming one of them, we are able to prove the other. Moreover, while Brouwer's fixed-point theorem proves the existence of a fixed point, it is a non-constructive result and does not give any insight as to how to find one. Indeed, the problem of finding a Brouwer fixed-point is proven to be PPAD-complete, a complexity class introduced by Papadimitriou et al [41], and is believed to be a difficult problem.

4.2.2 Kakutani's Fixed-Point Theorem

In this section, we present *Kakutani's fixed-point theorem*, which is a generalization of Brouwer's fixed-point theorem. Kakutani extended Brouwer's theorem in 1941 [42] to include set-valued functions. We begin with a few definitions, then state the theorem and provide an example in order to assist the reader in the theorem's comprehension.

Definition 4.10 (Set Valued Function)**.** A *set-valued function* ϕ from a set A to a set B is a rule that associates one or more points in B with each point in A. Formally it can be seen just as an ordinary function from A to the power set of B, written as $\phi : A \to 2^B$, such that $\phi(x)$ is non-empty for every $x \in A$.

Definition 4.11 (Closed Graph)**.** A set-valued function $\phi : A \to 2^B$ is said to have a *closed graph* if the set $C = \{(x, y) : y \in \phi(x)\}$ is a closed subset of the cartesian product $A \times B$.

We also extend our definition of a fixed point (Definition 4.7) to include fixed points of set-valued functions.

Definition 4.12 (Fixed Point of a Set-Valued Function)**.** Consider a set-valued function $\phi : A \to 2^A$. If there exists a point $x_0 \in A$ such that $x_0 \in \phi(x_0)$, then x_0 is called a *fixed point* of ϕ.

We are now ready to state Kakutani's fixed-point theorem.

Theorem 4.13. *Let S be a non-empty, compact and convex subset of a Euclidean space $\mathbb{R}^n$, and $\phi : S \to 2^S$ a set-valued function on S with a closed graph. Also, let $\phi(x)$ be non-empty and convex for all $x \in S$. Then, ϕ has at least one fixed point.*

Consider the following example to help with the comprehension of the theorem. Let $f(x)$ be a set-valued function defined on the closed interval $[0, 1]$ that maps a point $x \in [0, 1]$ to a subset of the closed interval $[1 - x/2,\ 1 - x/4]$. Then, f satisfies all the assumptions of Theorem 4.13, thus it has at least one fixed-point. If we plot the function on the closed interval $[0, 1]$ we get Figure 4.1.

Every point in the intersection of the red dotted line and the shaded grey area is a fixed point of f, which implies that, in this case, f has an infinite number of fixed points. Indeed, $x = 0.72$, denoted in Figure 4.1 with the dashed blue line, is a fixed point, since $[1 - 0.72/2, 1 - 0.72/4] = [0.64, 0.82]$ and $0.72 \in [0.64, 0.82]$.

Kakutani's fixed-point theorem has a significant number of applications to game theory. Specifically, as is discussed in Kakutani's original paper, the theorem can be used in zero-sum games, where each player's gain or loss in utility is exactly balanced

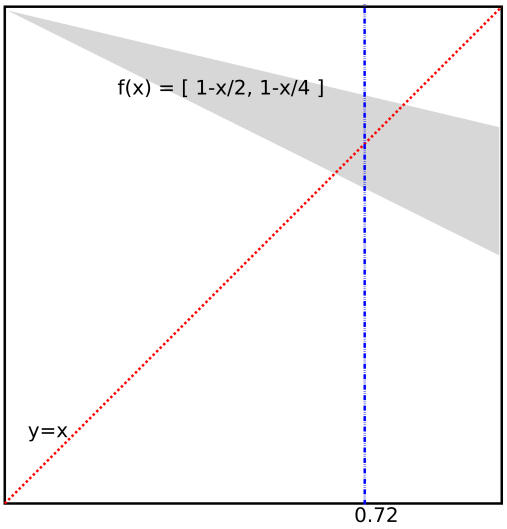

FIGURE 4.1: An example of Kakutani's fixed-point theorem

by the losses or gains in utility of the other players, to prove the minimax theorem. However, its most important contribution to game theory is perhaps its application in the proof by Nash of the existence of a mixed strategy Nash equilibrium in every finite game for any number of players, a work that later earned him a Nobel Prize in Economics [20].

In such a game, the tuples of mixed strategies chosen by each player constitutes the set S, and $\phi(\boldsymbol{x})$ is the function that, for the players' strategies in $\boldsymbol{x}$, returns a new tuple where each player's strategy is her best response to the other players' strategies in $\boldsymbol{x}$. It is possible for two or more strategies to be equally good, thus ϕ is set-valued. A Nash equilibrium of this game is defined as a fixed point of ϕ, specifically a tuple of strategies $\boldsymbol{x_0}$ where each player's strategy is a best response to the strategies of the other players in $\boldsymbol{x_0}$. The existence of such a fixed point, therefore a Nash equilibrium, follows directly from Kakutani's fixed-point theorem.

4.3 Concave Games

In this section we focus our attention on *concave games*. Concave games are characterized by two properties; every strategy lies inside a convex region of the product space of the individual strategies and each player's payoff function is concave. Before we continue, we define them formally below

Definition 4.14 (Concave Game)**.** A game with n players is called a *concave game* if and only if it satisfies the following properties:

- Every joint strategy $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, represented by a point in the product space of the individual strategy spaces, lies inside a convex and compact region R of the product space.
- Each player's payoff function ϕ_i is concave in his own strategy x_i.

Concave games have been studied a lot since they possess interesting properties that simplifies their analysis. Rosen, in his infamous theorem presented below, proved the existence of equilibrium points for every n-person concave game, and specified a certain property of the players' cost functions necessary for the equilibrium point to be unique. He also showed that if this property holds, the continuous best-response dynamics of the game converge to the unique equilibrium for any starting point.

We continue with the definition of a *socially concave game*, introduced by Even-Dar et al [43].

Definition 4.15 (Socially Concave Game)**.** A game with n players is called a *socially concave game* if and only if it satisfies the following properties:

- For every agent i, there exists a $\lambda_i > 0$ such that $f(\boldsymbol{x}) = \sum_i \lambda_i \phi_i(\boldsymbol{x})$ is concave in $\boldsymbol{x}$.
- For every agent i, the utility function $\phi_i(x_i, \boldsymbol{x}_{-i})$ is concave in x_i and convex in $\boldsymbol{x}_{-i}$.

If, in the definition above, we replace concavity with strict concavity, we define a *strict socially concave game*. In the next section, we show that strict socially concave games always satisfy the assumptions of Rosen's theorem, thus the continuous best-response dynamics of these games always converge to an equilibrium point for any starting point.

4.3.1 Rosen's Theorem

Certainly one of the most important results in Convex Optimization theory, Rosen's theorem sheds light on the convergence properties of n-person concave games. Proved by Rosen in 1965 [35], it consists of three parts, each one providing deep insight into the existence and uniqueness of Nash equilibria, and the necessary conditions for convergence of a concave game to them. We will present each section of Rosen's theorem as an individual theorem, without providing any proof since we prefer to present these ideas without getting into technical details.

Theorem 4.16. *For every n-person concave game, as defined in Definition 4.14, there exists at least one equilibrium point.*

This theorem follows directly from the application of Kakutani's fixed-point theorem (Section 4.2.2) on the convex and compact set of players' strategies R, along with a point-to-set mapping from R to R, that takes each point $\boldsymbol{x} \in R$ to the point where each player chooses her best-response to the strategy of all other players in $\boldsymbol{x}$, that maximizes her utility function.

While equilibrium points are guaranteed to exist in concave games by Theorem 4.16, their usefulness is limited since their uniqueness is not guaranteed. In fact, many games possess an infinite number of equilibrium points [44]. We continue by defining an additional concavity property that, when satisfied, guarantees the uniqueness of the equilibrium point.

Definition 4.17 (Diagonal Strict Concavity)**.** Consider a n-person concave game with the players' strategies lying inside a convex and compact set R, a coefficient vector $\boldsymbol{r} \in \mathbb{R}_{++}$ and a weighted nonnegative sum of the players' payoff functions

$$\sigma(\boldsymbol{x}, \boldsymbol{r}) = \sum_{i=1}^{n} r_i \phi_i(\boldsymbol{x}).$$

We also denote by $g(\boldsymbol{x}, \boldsymbol{r})$ the function

$$g(\boldsymbol{x}, \boldsymbol{r}) = \begin{bmatrix} r_1 \nabla_1 \phi_1(\boldsymbol{x}) \\ r_2 \nabla_2 \phi_2(\boldsymbol{x}) \\ \vdots \\ r_n \nabla_n \phi_n(\boldsymbol{x}) \end{bmatrix}$$

and call it the *pseudogradient* of $\sigma(\boldsymbol{x}, \boldsymbol{r})$. The function $\sigma(\boldsymbol{x}, \boldsymbol{r})$ is called *diagonally strictly concave* for $\boldsymbol{x} \in R$ and fixed $\boldsymbol{r}$, if the symmetric matrix $\left[G(\boldsymbol{x}, \boldsymbol{r}) + G^T(\boldsymbol{x}, \boldsymbol{r})\right]$ is negative definite for $\boldsymbol{x} \in R$, where $G(\boldsymbol{x}, \boldsymbol{r})$ is the Jacobian matrix of $g(\boldsymbol{x}, \boldsymbol{r})$.

Rosen utilizes the Karush-Kuhn-Tucker [45, 46] conditions to show that the diagonal scrict concavity of $\sigma(\boldsymbol{x}, \boldsymbol{r})$ is sufficient to prove the uniqueness of the equilibrium point.

Theorem 4.18. *Consider a n-person concave game that has at least one equilibrium point $\boldsymbol{x_0}$ from Theorem 4.16. Then, if $\sigma(\boldsymbol{x}, \boldsymbol{r})$ is diagonally strictly concave for some $\boldsymbol{r} \in \mathbb{R}_{++}$, the equilibrium point $\boldsymbol{x_0}$ is unique.*

Finally, we show that diagonal strict concavity on $\sigma(\boldsymbol{x}, \boldsymbol{r})$ implies the convergence of the continuous best-response dynamics of a n-person concave game to the unique equilibrium point, for any set of initial conditions. To prove this result, Rosen utilized his previous theorems along with the Karush-Kuhn-Tucker conditions to develop a gradient descent-like approach (Section 4.4) and prove that the distance of the current point

$\boldsymbol{x} \in R$ from $\boldsymbol{x_0}$ is decreasing with time, therefore it approaches zero asymptotically, and the system will converge asymptotically to $\boldsymbol{x_0}$.

Theorem 4.19. *Consider a n-person concave game, with the property that $\sigma(\boldsymbol{x}, \boldsymbol{r})$ is diagonally strictly concave for some $\boldsymbol{r} \in \mathbb{R}_{++}$. From Theorem 4.18, the game has a unique equilibrium point $\boldsymbol{x_0}$, and the continuous best-response dynamics of the game, where each player changes her strategy to one that maximizes her utility function given that all other players remain fixed, converge asymptotically to $\boldsymbol{x_0}$ for any initial point $\boldsymbol{x} \in R$.*

Strict socially concave games, as defined in Definition 4.15 with strict concavity, satisfy the conditions of Theorems 4.18 and 4.19, therefore, in strict socially concave games, the continuous best-response dynamics always converge to the unique equilibrium point for any starting point. However, if the conditions of Theorem 4.19 do not hold, the best-response dynamics need not converge to any equilibrium. For example, there exists a 2-person non-strict socially concave game in which the best-response dynamics do not converge. Furthermore, if players move towards the point chosen by best-response dynamics but not at a fixed proportional speed, the dynamics need not converge. In addition, in socially concave games, the sequential best-response dynamics, where each agent chooses her best response in turns, need not converge, even for the case of 2 players.

It should be noted here that Rosen's theorem was used by Bhawalkar et al [34] to prove the existence of a Nash equilibrium for the Generalized Asymmetric model (Section 3.5.2) and to prove that the Asymmetric k-NN model need not necessarily converge, since the agents' cost functions are not convex (Section 3.4.2).

4.4 Gradient Descent Methods

Up until now, we have analyzed several tools used to prove the existence of equilibrium points for opinion formation models. While some of them also provide a framework suitable to study the convergence properties of the models, they require strong assumptions and cannot be generalized properly. Here, we present one of the most significant methods used in optimization problems in order to locate a local (or global) minimum of a function, called *gradient descent*. Recall that an equilibrium point of an opinion formation model is always a local minimum of a potential function (Theorem 4.4), if our model admits to one. Therefore, we strive to minimize the potential function in order to reach the equilibrium point.

The gradient descent method makes use of the observation that the value of a function F decreases fastest if, from a specific point x, one takes steps proportional to the

negative of the gradient of F at x, $-\nabla F(x)$. Gradient descent, sometimes also called *steepest descent*, is a first-order technique so significant, it has spawned several first-order variations, which provide a framework for the analysis of models where simple gradient descent fails. However, all such variations fall under one of the two main categories of first-order techniques; gradient descent or *mirror descent*. Here, we present the ideas of both methods, without getting into many technical details, and show their usefulness in the analysis of opinion formation models.

4.4.1 Gradient Descent

Gradient descent, in its simplest form, is a first-order iterative algorithm used to locate a local minimum of a function.

Definition 4.20 (Gradient Descent)**.** Consider a multi-variable function F, and a point x_n. Let F be defined and differentiable in a neighborhood of x_n. Then, in the *gradient descent method*, at step $n+1$ of the algorithm, we move to the point

$$x_{n+1} = x_n - \gamma_n \nabla F(x_n),\ n \geq 0 \tag{4.6}$$

with γ_n a scalar value called the *step size* of the algorithm. Another way to write the above equation is the following

$$x_{n+1} = arg\min_x \left(\langle \nabla F(x_n), x \rangle + \frac{1}{\gamma_n} \|x - x_n\|_2^2 \right),\ n \geq 0 \tag{4.7}$$

where $\langle a, b \rangle$ is the inner product of a and b.

It is easy to see that for γ_n small enough, $F(x_{n+1}) \leq F(x_n)$. Therefore, we get the sequence $F(x_0) \geq F(x_1) \geq F(x_2) \geq \ldots$, which hopefully converges to a local minimum. Next, we see a set of conditions on F and γ that guarantee the convergence of the method to a minimum. Note that, for a convex function F, there is only a single global minimum, therefore, if the method converges, it will converge to the global minimum of F.

Theorem 4.21. *Let F be a convex function that is differentiable and its gradient ∇F is Lipschitz continuous. Also, consider a sequence of step sizes γ_n that satisfy the Wolfe conditions [47]. Then, there exists a point x^* such that $F(x^*) \leq F(x)$ for all x, and the gradient descent method defined =at Definition 4.20 converges to x^*.*

While the method of gradient descent is a fundamental tool of convex non-linear optimization, it is not without flaws. There are certain classes of pathological functions

that render gradient descent not a particularly useful technique. In addition, gradient descent does not converge very fast to the minimum, sometimes requiring many iterations to arrive at a specified distance from the point of convergence. Indeed, there are many techniques, based on Newton's method, that converge in fewer iterations. However, the computational cost of each iteration step is significantly higher in these techniques, and its simplicity and variety of applications make gradient descent almost always the first choice in non-linear optimization.

Gradient descent can be used in opinion dynamics to assist in the understanding of a model's convergence properties. It can provide insight into how the model's update rule will unfold over time and study whether its iteration will eventually converge. Indeed, gradient descent was used by Rosen in his proof of Theorem 4.19, and we can also show that in undirected linear models, like the Friedkin-Johnsel model (Section 2.1.2), the concurrent best-respone strategy used by all agents is equivalent to performing a gradient descent method on the model's potential function [48].

We proceed to show that the last claim holds. Consider an instance of the undirected FJ model. The game admits to a potential function

$$\Phi(\boldsymbol{x}) = \sum_{i=1}^{n} x_i \Big(w_{ii}(x_i - s_i) + \sum_{j \neq i}^{n} w_{ij}(x_i - x_j) \Big) \tag{4.8}$$

From (2.16), at each time step, each agent sets his opinion to

$$x_i(t+1) = w_{ii} s_i + \sum_{j \neq i}^{n} w_{ij} x_j(t) \tag{4.9}$$

or, in matrix form

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{s} \tag{4.10}$$

with $w_{ij} = 0$ for $j \notin \mathcal{N}_i$. Then, the potential function can be rewritten in matrix form as

$$\Phi(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{L}\boldsymbol{x} - \boldsymbol{B}\boldsymbol{x} \tag{4.11}$$

where $\boldsymbol{L}$ is the Laplacian matrix of the model, with $L_{ii} = \sum_{j=1}^{n} w_{ij}$ and $L_{ij} = -w_{ij}$ for all $j \neq i$. Since our model is undirected, $\boldsymbol{L}$ is a symmetric positive semidefinite matrix, thus Φ is a quadratic and convex function of $\boldsymbol{x}$. We can calculate the gradient of Φ

$$\nabla\Phi(\boldsymbol{x}) = \boldsymbol{L}\boldsymbol{x} - \boldsymbol{B} \tag{4.12}$$

Since we assume normalized weights, $\sum_{j=1}^{n} w_{ij} = 1$ for all agents i, we can rework (4.9) to simulate a gradient descent method

$$\boldsymbol{x}(t+1) = \boldsymbol{x}(t) - \nabla\Phi(\boldsymbol{x}(t)) \tag{4.13}$$

with $\boldsymbol{x}(0) = \boldsymbol{s}$ in the FJ model. Next, we will prove that the gradient descent method above converges to the equilibrium $\boldsymbol{x}^* = \boldsymbol{L}^{-1}\boldsymbol{B}$. To do that, we first need to prove the following lemmas

Lemma 4.22. *For any time step $t \geq 0$, we have*

$$\boldsymbol{x}(t+1) - \boldsymbol{x}^* = \boldsymbol{A}(\boldsymbol{x}(t) - \boldsymbol{x}^*) \tag{4.14}$$

where $\boldsymbol{A}$ is the averaging matrix in (4.10).

Proof. Substituting (4.12) into (4.13) gives us

$$\boldsymbol{x}(t+1) = \boldsymbol{x}(t) - \boldsymbol{L}\boldsymbol{x}(t) + \boldsymbol{B}$$

But $\boldsymbol{B} = \boldsymbol{L}\boldsymbol{x}^*$, therefore

$$\begin{aligned}\boldsymbol{x}(t+1) &= \boldsymbol{x}(t) - \boldsymbol{L}\boldsymbol{x}(t) + \boldsymbol{L}\boldsymbol{x}^* \\ &= \boldsymbol{x}(t) - \boldsymbol{L}(\boldsymbol{x}(t) - \boldsymbol{x}^*)\end{aligned}$$

Subtracting $\boldsymbol{x}^*$ from the equation above gives us

$$\begin{aligned}\boldsymbol{x}(t+1) - \boldsymbol{x}^* &= \boldsymbol{x}(t) - \boldsymbol{x}^* - \boldsymbol{L}\boldsymbol{x}(t) + \boldsymbol{L}\boldsymbol{x}^* \\ &= (\boldsymbol{I} - \boldsymbol{L})(\boldsymbol{x}(t) - \boldsymbol{x}^*)\end{aligned}$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix. We observe that $\boldsymbol{I} - \boldsymbol{L} = \boldsymbol{A}$, where $\boldsymbol{A}$ is the averaging matrix in (4.10), therefore

$$\boldsymbol{x}(t+1) - \boldsymbol{x}^* = \boldsymbol{A}(\boldsymbol{x}(t) - \boldsymbol{x}^*)$$

□

For simplicity, let $\boldsymbol{e}_t = \boldsymbol{x}(t) - \boldsymbol{x}^*$. Next, we show that if $\boldsymbol{A}$'s eigenvalues lie in $(-1, 1)$, then the distance for the equilibrium, $\boldsymbol{e}_t$ strictly decreases at each time step t.

Lemma 4.23. *If* $\max_i |\lambda_i(\boldsymbol{A})| < 1$*, then, for any time step* $t \geq 0$*, we have*

$$\|\boldsymbol{e}_{t+1}\|_2^2 < \|\boldsymbol{e}_t\|_2^2 \tag{4.15}$$

Proof. Utilizing Lemma 4.22, we have that

$$\begin{aligned}
\|\boldsymbol{e}_t\|_2^2 - \|\boldsymbol{e}_{t+1}\|_2^2 &= \|\boldsymbol{e}_t\|_2^2 - \boldsymbol{e}_{t+1}^T \boldsymbol{e}_{t+1} \\
&= \|\boldsymbol{e}_t\|_2^2 - (\boldsymbol{A}\boldsymbol{e}_t)^T (\boldsymbol{A}\boldsymbol{e}_t) \\
&= \boldsymbol{e}_t^T \boldsymbol{e}_t - \boldsymbol{e}_t^T \boldsymbol{A}^2 \boldsymbol{e}_t \\
&= \boldsymbol{e}_t^T (\boldsymbol{I} - \boldsymbol{A}^2) \boldsymbol{e}_t
\end{aligned}$$

All eigenvalues of $\boldsymbol{A}$ lie in $(-1, 1)$, thus all eigenvalues of $\boldsymbol{A}^2$ lie in $(0, 1)$. Therefore, all eigenvalues of $\boldsymbol{I} - \boldsymbol{A}^2$ are strictly positive, thus $\boldsymbol{I} - \boldsymbol{A}^2$ is a positive definite matrix, which implies that, for any $\boldsymbol{e}_t$, $\boldsymbol{e}_{t+1}$

$$\|\boldsymbol{e}_t\|_2^2 - \|\boldsymbol{e}_{t+1}\|_2^2 > 0$$

□

We can now combine these two lemmas to prove the following theorem

Theorem 4.24. *The gradient descent method described at (4.13) converges to the Nash equilibrium* $\boldsymbol{x}^*$ *of the FJ model, which is also the point where the potential function* Φ *attains its minimum value.*

Proof. Recall that, in the FJ model, there exists at least one agent i that has $w_{ii} > 0$. Therefore, $\boldsymbol{A}$ is a substochastic matrix and its maximum eigenvalue $\max_i |\lambda_i(\boldsymbol{A})| < 1$. From Lemmas 4.22 and 4.23, the gradient descent method described at (4.13) decreases the distance from $\boldsymbol{x}^*$ at each iteration step, therefore it converges to the Nash equilibrium $\boldsymbol{x}^*$ of the FJ model. □

Therefore, we observe that, in undirected linear models, concurrent best-response is equivalent to performing a gradient descent method on the model's potential function, and in these models all agents converge to a stable state.

4.4.2 Mirror Descent

In an effort to generalize the gradient descent method beyond Euclidean metric spaces, a variant of the method called *mirror descent* was developed. This method utilizes the concept of *Bregman divergence* instead of the Euclidean norm as a measure of displacement, which we define below

Definition 4.25 (Bregman Divergence)**.** Let $\psi : \Omega \to \mathbb{R}$ be continuously differentiable and strictly convex function, defined on a closed convex set Ω. Then, for any two points $x, y \in \Omega$, the *Bregman divergence* under ψ is defined as

$$\mathcal{D}_\psi(x, y) = \psi(x) - \psi(y) - \langle \nabla\psi(y), x - y\rangle \tag{4.16}$$

For example, to get the Euclidean distance, we have $\psi(x) = \|x\|_2^2/2$. Then, $\mathcal{D}_\psi(x, y) = \|x - y\|_2^2/2$. Intuitively, the Bregman divergence calculates the difference between the value of ψ at x and the first order Taylor expansion of ψ around y, evaluated at x. Bregman divergence generalizes the squared Euclidean distance to a class of distances, all sharing similar properties, and has numerous applications in machine learning and clustering. It posseses several useful properties, and we present a few of them below

- Nonnegativity: $\mathcal{D}_\psi(x, y) \geq 0$ for all x, y. Specifically, $\mathcal{D}_\psi(x, y) = 0$ if and only if $x = y$.
- Asymmetry: In general, we have that $\mathcal{D}_\psi(x, y) \neq \mathcal{D}_\psi(y, x)$.
- Linearity in ψ: For any $a > 0$, $\mathcal{D}_{\psi+a\phi}(x, y) = \mathcal{D}_\psi(x, y) + a\mathcal{D}_\phi(x, y)$.

Now we are ready to define the method of mirror descent

Definition 4.26 (Mirror Descent)**.** Consider a multi-variable function F, a point x_n and a function ψ that is continuously differentiable and strictly convex. Let F be defined and differentiable in a neighborhood of x_n. Then, in the *mirror descent method*, at step $n + 1$ of the algorithm, we move to the point

$$x_{n+1} = arg\min_x \Big(\langle \nabla F(x_n), x\rangle + \frac{1}{\gamma_n}\mathcal{D}_\psi(x, x_n)\Big),\ n \geq 0 \tag{4.17}$$

Notice that, since for $\psi(x) = \|x\|_2^2/2$ the Bregman divergence coincides with the Euclidean distance, for this function ψ the mirror descent method coincides with the gradient descent, as defined in Definition 4.20. While mirror descent has not seen broad use in the field of opinion dynamics as of yet, we believe that it can be utilized to provide significant insight on the convergence properties of several models.

4.5 Energy as a Generating Function

We conclude our presentation of fundamental techniques used to analyze opinion formation models with the concept of the *energy of a system*. The energy approach was developed as a way to study the convergence properties of complex models, where step-by-step methods fail. In several systems, we may lack the guarantees about specific quantities changing between two consecutive time steps. However, in these cases, we can use the concept of the system's energy as a generating function that bands together several continuous time steps in order to observe a specific change in a property of the system. This change, for example, could be a decrease in the distance from an equilibrium point or a decrease in a potential-like function of the model, if one exists.

The notion of a system's energy was first introduced by Chazelle [33] to study a generalization of opinion formation models, called *influence systems*. Therefore, the technique, besides powerful, is also quite general and possibly applicable on numerous models. Consequently, in our opinion, attempts to apply the ideas presented here on several variations of the Hegselmann-Krause model and other non-linear models in general, would hold significant merit and may provide fascinating new results.

4.5.1 Definition of the Total s-Energy

We begin by defining the concept of the *total s-energy* of a multiagent system, introduced by Chazelle, who also proved that its convergence for any real $s > 0$. [33].

Definition 4.27 (Total s-Energy)**.** Consider an infinite sequence of graphs G_0, G_1, G_2, ..., where each G_t has n nodes labeled $1, 2, \ldots, n$, with each node representing an agent. We assume the agents' opinions lie in a d-dimensional Euclidean space and denote agent i's opinion at time t with $x_i(t) \in \mathbb{R}^d$. Then, the *total s-energy* of this system is defined as

$$E(s) = \sum_{t \geq 0} \sum_{(i,j) \in G_t} \|x_i(t) - x_j(t)\|_2^s \tag{4.18}$$

where the exponent $s \in \mathbb{C}$ is, in the most general setting, a complex variable.

We observe that $E(s)$ encodes all of the edge lengths for every G_t in the graph sequence. We call this graph sequence that shares the same nodes the *communication network* of the system, and we make no assumptions about it in our definition above. In fact, this model is so general, there is no obvious reason why $E(s)$ should ever converge, for any s. Later in this chapter we provide a proof that $E(s)$ converges for any $s \in \mathbb{R}_+$. However, before we continue, we provide some general intuition behind the concept of

s-energy, in order to assist in the understanding of the ideas and results presented in this section.

4.5.2 Bidirectional Systems

Recall the constraints imposed on the communication network of the Network-HK model presented in Section 3.1.1. We attempt to generalize them by defining *bidirectional agreement systems* in general. Our definition introduces the model for the 1-dimensional case, however it contains all of the necessary ideas and the extension of the model to higher dimensions can be done in many ways in a straightforward fashion.

Definition 4.28 (Bidirectional Agreement System)**.** Consider n agents expressing opinions $x_1(t), x_2(t), \ldots, x_n(t) \in \mathbb{R}$ at time t. The communication network consists of an infinite sequence of graphs $G_0, G_1, G_2, \ldots$, with G_t being a function of the system's configuration at times $0, \ldots, t-1$. Let $\mathcal{N}_i(t) = \{j : (i,j) \in G\}$ denote the set of neighbors of i, $m_i(t) = \min_{j \in \mathcal{N}_i(t)} x_j(t)$ the minimum and $M_i(t) = \max_{j \in \mathcal{N}_i(t)} x_j(t)$ the maximum opinion in i's neighborhood at time t. Also, let $0 < \rho \leq 1/2$ be an *agreement parameter* that is time-invariant and uniform for all agents. Then, in a *bidirectional agreement system*, at time t each agent i moves to $x_i(t+1)$, where

$$(1-\rho)m_i(t) + \rho M_i(t) \leq x_i(t+1) \leq \rho m_i(t) + (1-\rho)M_i(t) \tag{4.19}$$

Note that the model does not make any assumptions on the generation of each G_t, nor on their connectivity properties. It is believed that bidirectional systems are the widest class of systems that allow for reasoning on their convergence properties, since the general case of directed graphs precludes such analysis. In addition, note that the model is nondeterministic, with the communication network and the agents' motion being completely arbitrary, and it does not imply symmetry among neighbors. Indeed, the set of constraints that Definition 4.28 imposes on agent behavior is fairly weaker than the usual set of constraints associated with bidirectional models.

Recall that in non-linear systems, $\boldsymbol{x}(t+1) = \boldsymbol{A}(t)\boldsymbol{x}(t)$, where $\boldsymbol{A}$ is a row-stochastic matrix with positive entries $a_{ij}(t)$ for all i, j where $j \in \mathcal{N}_i(t)$. Then, if we denote by l_i the leftmost and by r_i the rightmost neighbor of i, we get a set of two constraints, which follow directly from (4.19)

- Mutual confidence: Pairs of a_{ij}, a_{ji} do not have exactly one zero; they are either both positive or both zero.
- No extreme influence: For any agent i that is not fully-stubborn, $\max\{a_{il_i}(t), a_{ir_i}(t)\} \leq 1 - \rho$.

We immediately see that the above conditions are weaker than those of Section 3.1.1. Intuitively, for bidirectional systems to converge asymptotically, agents may be influenced a lot by non-extreme positions but must be the influence that extreme positions exert upon them must be bounded. It has been shown that such bidirectional systems converge asymptotically [24, 25, 49], and the convergence rate is bounded by $\rho^{-\mathcal{O}(n)}$ [33].

4.5.3 Bounds on the Total s-Energy

It is easy to see that understanding opinion formation models, or agreement systems, in their most general setting is equivalent to understanding backward products of stochastic matrices

$$\boldsymbol{A}(t)\boldsymbol{A}(t-1)\dots\boldsymbol{A}(0)$$

which relate to time-inhomogeneous Markov chains. Although not much is known about such Markov chains, we can make some interesting observations here. Note that if a product such as the one above converges, then as time goes by, the product will tend to a matrix of rank one.

To see why this is true, consider the following geometric interpretation. Each row of $\boldsymbol{A}(0)$ corresponds to a specific point in $\mathbb{R}^n$. Construct a convex polytope equal to the convex hull of all these points, denoted by $conv(\boldsymbol{A}(0))$. When $\boldsymbol{A}(0)$ is multiplied by $\boldsymbol{A}(1)$, each row of the product is a convex combination of the rows of $\boldsymbol{A}(0)$, hence each point of the product is a convex combination of the points specified by $\boldsymbol{A}(0)$ and lies inside the convex hull $conv(\boldsymbol{A}(0))$. Therefore, $conv(\boldsymbol{A}(1)\boldsymbol{A}(0)) \subseteq conv(\boldsymbol{A}(0))$. Repeating the process, we get a sequence of convex polytopes

$$conv(\boldsymbol{A}(t)\dots\boldsymbol{A}(0)) \subseteq \dots \subseteq conv(\boldsymbol{A}(1)\boldsymbol{A}(0)) \subseteq conv(\boldsymbol{A}(0))$$

that is decreasing in volume. Recall Definition 3.3 of the coefficient of ergodicity and observe that it essentially is a measure of how fast this sequence of convex polytopes decreases in volume, and the matrices approach a matrix of rank one. However, it is a local tool in the sense that it analyzes the model in a step-by-step fashion. In contrast, the notion of the total s-energy presented in this section is a global tool; it monitors the decrease of this sequence over all time steps, with parameter s essentially playing the role of frequency in Fourier analysis.

As stated before, there is no obvious reason as to why the total s-energy of a system should ever converge. However, Chazelle proved in 2010 that, for bidirectional agreement systems where all agents express opinions in $[0,1]^d$, the total s-energy is bounded for any

real $s > 0$ and also provided an upper bound on the convergence rate [33]. We present these two theorems here for the 1-dimensional case, and we refer to Chazelle's paper for the proofs.

Theorem 4.29. *Consider a n-agent bidirectional agreement system where each agent expresses an opinion in* $[0, 1]$*, and let* $E_n(s)$ *denote the maximum value of the total s-energy, over all times and all n-node graph sequences. Then,*

$$E_n(s) \leq \begin{cases} \rho^{-\mathcal{O}(n)} & for\ \ s = 1. \\ s^{1-n}\rho^{-n^2-\mathcal{O}(1)} & for\ \ 0 < s < 1 \end{cases} \tag{4.20}$$

Since no edge length exceeds 1, $E_n(s) \leq E_n(1)$ *for* $s \geq 1$*, therefore* $E_n(s)$ *is bounded from above for any real* $s > 0$*.*

However, if we attempt to bound the convergence rate of the system, we are immediately faced with the obvious difficulty that, in an adversarial setting, an adversary can always provide us with a graph G_t that is an independent set for as long as it wants and then at some time into the future connect all edges permanently to the network in order to make the agents reach consensus. We circumvent this difficulty by defining notion of asymptotic convergence. Specifically, given $0 < \varepsilon < 1/2$, we say that step t is *trivial* if all the edges in G_t have length at most ε. We then proceed to bound the *communication count* C_ε, defined as the total number of non-trivial steps for a given ε. Intuitively, this notion of convergence ignores microscopic motions of the agents and specifies defines convergence as the event where all agents have formed independent clusters of radius ε. From a macroscopic point of view, this notion implies that the system eventually freezes.

Theorem 4.30. *Consider a n-agent bidirectional agreement system where each agent expresses an opinion in* $[0, 1]$*. Then, the maximum communication count is bounded by above*

$$C_\varepsilon \leq \min\Big\{\frac{1}{\varepsilon}\rho^{-\mathcal{O}(n)}, (log\frac{1}{\varepsilon})^{n-1}\rho^{-n^2-\mathcal{O}(1)}\Big\} \tag{4.21}$$

4.5.4 Kinetic s-Energy

While the concept of the total s-energy is extremely useful for macroscopic analysis of agreement systems, sometimes variants of the concept are more convenient, and such is the case with opinion formation models. In particular, we define the *kinetic s-energy* of a system as follows

Definition 4.31 (Kinetic s-Energy)**.** Consider an infinite sequence of graphs G_0, G_1, G_2, ..., where each G_t has n nodes labeled $1, 2, \ldots, n$, with each node representing an agent. We assume the agents' opinions lie in a d-dimensional Euclidean space and denote agent i's opinion at time t with $x_i(t) \in \mathbb{R}^d$. Then, the *kinetic s-energy* of this system is defined as

$$K(s) = \sum_{t \geq 0} \sum_{i=1}^{n} \|x_i(t+1) - x_i(t)\|_2^s \tag{4.22}$$

Recall our use of the kinetic s-energy for $s = 2$ in Section 3.3, to prove the convergence of the HK model with fully-stubborn agents. Such applications of the concept demonstrate the usefulness of this mathematical tool in the analysis of opinion formation models and their convergence properties. We believe that, when applied correctly, the concept of the kinetic s-energy can simplify and, most importantly, generalize existent proofs of convergence on several models, as well as provide new insight on the convergence properties of models we know very little about.

Chapter 5

Convergence of Variations of the Hegselmann - Krause Model

In this chapter, we analyze the Network-HK and Inertial-HK models using the concept of the system's s-Energy (Section 4.5). Specifically, we utilize the system's kinetic s-energy to prove that the Network-HK model converges to an equilibrium, as stated by Theorem 3.5. Furthermore, we wish to provide a deeper understanding of the message-passing protocol introduced in the Inertial-HK model (Section 3.3.2), as we believe its conception is not trivial and there is significant merit in understanding the thought process of deriving such a protocol, if one hopes to apply the s-energy approach to study the convergence properties of other models as well.

5.1 Energy Approach to the Network-HK Model

The purpose of this section is to utilize the s-energy approach to prove Theorem 3.5. Recall that, to show that Theorem 3.5 holds, we require only Lemmas 3.2 and 3.4, namely that there is a time t when the agents have split up into weakly connected sets, and that every weakly connected set of agents eventually arrives at consensus.

Our proof of Lemma 3.2 is a simple temporal graph argument that we retain in this proof. Therefore, we focus our analysis on Lemma 3.4. Our proof is quite similar to the proof of Theorem 3.12 which shows that the Inertial-HK model converges to a stable state. As in that case, we will first provide an upper bound of the kinetic s-energy of the system, for the case $s = 2$, and then proceed to show that when the kinetic 2-energy is bounded, any weakly connected set of agents converges to a single opinion.

Lemma 5.1. *Let $(G(V, E), \varepsilon, \boldsymbol{x}(0))$ be an instance of the Network-HK model, where V is weakly connected. Then, all agents converge to a single opinion x^*.*

Proof. Note that the bound on the kinetic 2-energy of the Inertial-HK system makes no assumptions on the connectivity of agents. Therefore, we can use the same message-passing protocol to arrive at exactly the same bound for the kinetic 2-energy, with very few alterations on the proof. Specifically, agent i's neighborhood is different in the Network-HK model, as it is influenced by the underlying graph, and is equal to

$$\mathcal{N}_i(G_t, t, \varepsilon) = \{j : \{i, j\} \in E \text{ and } |x_i(t-1) - x_j(t-1)| \leq \varepsilon\} \tag{5.1}$$

Furthermore, since the Network-HK model extends the original HK model simply with the addition of an underlying graph, all agents' inertias are equal to 1. Thus, for the kinetic 2-energy, we get

$$K(2) \leq n^2/4 \tag{5.2}$$

It is easy to see that the proof of Lemma 3.11 holds even if we make the alterations above, as they do not have any effect on the message-passing protocol or the arguments used. However, the change in the model, and specifically in the neighborhood of each agent, renders the proof of Theorem 3.12 insufficient. Therefore, we need to present a different argument in order to prove that, in the Network-HK model, a weakly connected set of agents reaches consensus, when the kinetic 2-energy is bounded.

The upper bound on the kinetic 2-energy that we get by 5.2 shows that, for any arbitrarily small $\delta > 0$, there exists a time step t_δ such that no agent moves by a distance of more than δ at any time $t \geq t_\delta$. Consider a fixed time $t_0 > t_\delta$, and let, for brevity, $x_i = x_i(t_0)$ and $\mathcal{N}_i = \mathcal{N}_i(G_{t_0}, t_0, \varepsilon)$ for each agent i. Again, we use primes and double primes to indicate the equivalent quantities for times $t_0 + 1$ and $t_0 + 2$.

Recall our notation for the four distinct sets of agents from Section 3.3.2

- Let L_i^{in} be the set of agents located at $x_i - \varepsilon - \mathcal{O}(\delta)$ at time t_0 and at $x_i - \varepsilon + \mathcal{O}(\delta)$ at time $t_0 + 1$. This set consists of the agents that joined i's neighborhood at $t_0 + 1$ from the left side of agent i.
- Let L_i^{out} be the set of agents located at $x_i - \varepsilon + \mathcal{O}(\delta)$ at time t_0 and at $x_i - \varepsilon - \mathcal{O}(\delta)$ at time $t_0 + 1$. This set consists of the agents that left i's neighborhood at $t_0 + 1$ from the left side of agent i.
- Let R_i^{in} be the set of agents located at $x_i + \varepsilon + \mathcal{O}(\delta)$ at time t_0 and at $x_i + \varepsilon - \mathcal{O}(\delta)$ at time $t_0 + 1$. This set consists of the agents that joined i's neighborhood at $t_0 + 1$ from the right side of agent i.

- Let R_i^{out} be the set of agents located at $x_i+\varepsilon-\mathcal{O}(\delta)$ at time t_0 and at $x_i+\varepsilon+\mathcal{O}(\delta)$ at time t_0+1. This set consists of the agents that left i's neighborhood at t_0+1 from the right side of agent i.

It is obvious that all the sets introduced above are disjoint, and their union is the symmetric difference between $\mathcal{N}_i$ and $\mathcal{N}_i'$. The locations x_i' and x_i'' of agent i at times t_0+1 and t_0+2 are given by

$$|\mathcal{N}_i|x_i' = \sum_{j\in\mathcal{N}_i\cap\mathcal{N}_i'} x_j + \sum_{j\in L_i^{out}\cup R_i^{out}} x_j$$
$$|\mathcal{N}_i'|x_i'' = \sum_{j\in\mathcal{N}_i\cap\mathcal{N}_i'} x_j' + \sum_{j\in L_i^{in}\cup R_i^{in}} x_j'$$

Since all x_k' and x_k'' are of the form $x_k \pm \mathcal{O}(\delta)$, subtracting the two identities above shows that

$$(|\mathcal{N}_i'| - |\mathcal{N}_i|)x_i = (|L_i^{in}| - |L_i^{out}|)(x_i-\varepsilon) + (|R_i^{in}| - |R_i^{out}|)(x_i+\varepsilon) \pm \mathcal{O}(\delta)n \tag{5.3}$$

The dynamics are translation-invariant, thus we can set $x_i = 0$. If we choose a small enough δ, the integrality of the set cardinalities implies that the net flow of neighbors on the left of agent i is the same as it is on the right

$$|L_i^{out}| - |L_i^{in}| = |R_i^{out}| - |R_i^{in}| \tag{5.4}$$

As with the proof in Section 3.3.2, our goal is to show that, at time t_0, no agent is undergoing a change of neighbors, thus all four terms of (5.4) are equal to zero. Suppose that at least one term of (5.4) is positive, to arrive at a contradiction. Then, we focus on the agents that are undergoing a change of neighbors between times t_0 and t_0+1. Among these agents, we choose the one that ends up the furthest to the right at time t_0+1, breaking ties by picking the agent with the largest index. We call this agent i, and, without loss of generality, we assume that $x_i(t_0+1) \geq x_i(t_0)$, which means that agent i moves to the right at time t_0+1. Later on, we show why our argument holds even if $x_i(t_0+1) < x_i(t_0)$ and agent i moves to the left at time t_0+1.

First of all, we see that R_i^{out} must be empty, since if an agent $j \in R_i^{out}$, that would imply the existence of an agent undergoing a change of neighbors and landing to the right of agent i, which contradicts our definition of i. Therefore, $R_i^{out} = \emptyset$, which, in

turn, implies that $L_i^{in} \neq \emptyset$, since we assumed that not all four terms of (5.4) can be zero. We proceed to pick an agent $k \in L_i^{in}$, and we study his motion between times t_0 and $t_0 + 1$. Note that, since agent k joins i's neighborhood and i moves to the right at time $t_0 + 1$, k must move to the right as well, thus we have $x_k(t_0 + 1) \geq x_k(t_0)$. We distinguish between three cases

i. $|R_k^{out}| > |R_k^{in}|$: This means that there are more agents which leave k's neighborhood from the right than join it. However, our set of agents is weakly connected. Therefore, there exists a time t in the future where each of these agents has to reenter k's neighborhood. Since R_k is finite, it follows that there is a time where there cannot be more agents leaving R_k than joining and, at that time, $|R_k^{out}| \leq |R_k^{in}|$. Thus, this case is reduced to either $|R_k^{out}| < |R_k^{in}|$, or $|R_k^{out}| = |R_k^{in}|$.

ii. $|R_k^{out}| < |R_k^{in}|$: This implies that $L_k^{in} \neq \emptyset$. Therefore, we pick an agent in L_k^{in} and continue inductively, noting that every time this case holds, we pick an agent that is to the left of all other agents picked so far. Because the set of agents is finite, in this case we eventually pick the leftmost agent l which has $L_l^{in} = \emptyset$ by definition, thus arriving at a contradiction.

iii. $|R_k^{out}| = |R_k^{in}|$: Here, we can distinguish even further between the case where $L_k^{in} \neq \emptyset$ and the case where $L_k^{in} = \emptyset$. In the first, we can pick an agent in L_k^{in} as the case above, and eventually arrive at a contradiction. In the second, we have $L_k^{in} = \emptyset$, $|R_k^{out}| = |R_k^{in}| > 0$, which imply $L_k^{out} = \emptyset$. We observe that, since the other two cases are resolved, the only case that remains is the case where $|R_k^{out}| = |R_k^{in}|$, continuously, for all times $t \geq t_0$. Therefore, in this case we see that it is possible for the communication graph to not stabilize, as edges can appear and dissapear arbitrarily, as long as agent i has an edge with any agent to his right sometime in the future, to preserve the property of weak connectivity. Thus, we arrive at an important observation; the communication graph need not converge for the agents to reach consensus. Indeed, since $|R_k^{out}| = |R_k^{in}| > 0$ and $L_k^{in} = L_k^{out} = \emptyset$, we note that agent k at every time step has at most $\frac{n-1}{2}$ agents to the right, and for any one of these agents, say j, we have that $x_j(t) - x_k(t) \geq \varepsilon - \delta$. Thus, at every time step, agent k moves by a fraction of at least $\frac{2}{n-1}(\varepsilon - \delta)$ to the right. But every agent moves at most δ, therefore, for $\delta < \frac{2\varepsilon}{n+1}$ we arrive once again at a contradiction.

It is easy to see that our argument does not use time-directionality, except for the case $|R_k^{out}| > |R_k^{in}|$. However, in this case, our argument still holds if we reverse the direction of time, since we can switch the arguments for the first two cases and have an argument that holds for the reverse direction of time. Thus, in the case where agent i

moves to the left at time $t_0 + 1$ and $x_k(t_0 + 1) < x_k(t_0)$, we can reverse the direction of time, exchange the roles of t_0 and $t_0 + 1$, get $x_k(t_0 + 1) \geq x_k(t_0)$ and our argument above still holds. Note that we must swap the superscripts *in* and *out* of the sets and, by symmetry, we now have to pick the agent that starts, rather than ends up, furthest to the right. By the preservation of the agents' ordering of the HK model, we get that this change makes no difference to our argument.

The finiteness of the set of agents V implies that, eventually, we arrive at a contradiction, through either one of these cases. Therefore, our assumption that not all four terms of (5.4) can be zero is wrong, and there is a time step after which all agents are endowed with a fixed set of neighbors. Thus, the communication network converges, and the model's dynamics are specified by the powers of a fixed stochastic matrix $\boldsymbol{A}$. Therefore, we essentially have an instance of the DeGroot model (Section 2.1.1), enhanced with self-loops which imply that $\boldsymbol{A}$'s diagonal is strictly poistive. As we have seen in Section 2.1.1, in this model all agents converge to a single opinion x^*, and our proof is complete. □

Now we can combine Lemma 3.2 and Lemma 5.1 to show, once again, that Theorem 3.5 holds, thus proving, with the use of the s-energy of the system, that the Network-HK model converges to a stable state.

5.2 Analysis of the Inertial-HK Model

In this section we attempt to shed light on the convergence proof of the Inertial-HK model (Section 3.3.2). Specifically, we focus on Lemma 3.11, which provides an upper bound on the kinetic 2-energy of the system. We acknowledge that the message-passing protocol presented in that proof seems a little arbitrary at first, and there seems to be no intuition as to why agent i spends and exchanges the amounts she does. We provide a different line of thinking and work the model backwards to arrive, surprisingly, at exactly the message-passing protocol that we demonstrated earlier. This contribution follows from our wish to provide some much wanted intuition behind the proof of Section 3.3.2, and assist the reader in its understanding.

First, we rework the definition of the kinetic 2-energy and tailor it to our model, as this will assist our thought process. In the rest of this section, $K(2)$ will denote the kinetic 2-energy of the system between two time steps t and $t + 1$, for convenience. Recall the definition of the kinetic 2-energy (Definition 4.31) and of the Inertial-HK model (3.7). If we consider two time steps t and $t + 1$ and focus only on a single agent i, combining these two equations, we get

$$K_i(2) = \|x_i(t+1) - x_i(t)\|_2^2 \tag{5.5}$$

$$= \left\| \lambda_i(t) x_i(t) + \frac{\lambda_i(t)}{|\mathcal{N}_i(t)|} \sum_{j \in \mathcal{N}_i(t)} x_j(t) \right\|_2^2 \tag{5.6}$$

$$= \frac{\lambda_i^2(t)}{|\mathcal{N}_i(t)|^2} \left\| \sum_{j \in \mathcal{N}_i(t)} x_j(t) - x_i(t) \right\|_2^2 \tag{5.7}$$

We note that all $\lambda_i(t)$ are bounded by 1 and $|\mathcal{N}_i(t)| \geq 1$, as $i \in \mathcal{N}_i(t)$, for all agents i and times t. Utilizing these observations, as well as the triangle inequality, we arrive at

$$K_i(2) \leq \sum_{j \in \mathcal{N}_i(t)} \|x_j(t) - x_i(t)\|_2^2$$

$$K_i(2) \leq \sum_{j \in \mathcal{N}_i(t)} \|d_{ij}\|_2^2 \tag{5.8}$$

where $d_{ij} = x_i(t) - x_j(t)$.

In our opinion, to arrive smoothly at the protocol demonstrated earlier, we must work backwards in the model, and first address the important question; when we say say we want to bound the kinetic 2-energy, what exactly do we mean? It is obvious that if we can derive a function $f(\boldsymbol{x}(t))$ such that $f(\boldsymbol{x}(t)) > 0$ and $f(\boldsymbol{x}(t)) \geq K(2) + f(\boldsymbol{x}(t+1))$ for any times t and $t+1$, then the kinetic 2-energy is bounded by $f(\boldsymbol{x}(0))$ and, if that quantity is finite, our work is done. This function plays the role of $C(\boldsymbol{x}(t))$ in the previous proof, in other words it denotes the amount of "money" that each agent possesses at time t. Thus, we need to "guess" a function f with the property that, for all times t and $t+1$,

$$f(\boldsymbol{x}(t)) \geq \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i(t)} \|d_{ij}\|_2^2 + f(\boldsymbol{x}(t+1))$$

Breaking up f into parts f_i for each agent i, where $f(\boldsymbol{x}(t)) = \sum_{i=1}^{n} f_i(\boldsymbol{x}(t))$, gives us

$$f_i(\boldsymbol{x}(t)) \geq \sum_{j \in \mathcal{N}_i(t)} \|d_{ij}\|_2^2 + f_i(\boldsymbol{x}(t+1)) \tag{5.9}$$

The reasonable line of thinking here is to let $f_i(\boldsymbol{x}(t+1)) = 0$, and assume the lowest possible value for $f_i(\boldsymbol{x}(t))$, therefore arriving at a guess

$$f_i(\boldsymbol{x}(t)) = \sum_{j \in \mathcal{N}_i(t)} \|d_{ij}\|_2^2 \tag{5.10}$$

Now, if we define $d'_{ij} = x_i(t+1) - x_j(t+1)$, $\mathcal{N}_i = \mathcal{N}_i(t)$ and $\mathcal{N}'_i = \mathcal{N}_i(t+1)$, we can observe the flow of "money" of agent i between times t and $t+1$ as

$$f_i(\boldsymbol{x}(t+1)) - f_i(\boldsymbol{x}(t)) = \sum_{j \in \mathcal{N}'_i} \|d'_{ij}\|_2^2 - \sum_{j \in \mathcal{N}_i} \|d_{ij}\|_2^2 \tag{5.11}$$

The equation above demonstrates how much money i has to spend for each agent. At this point, we need to look closer at $\mathcal{N}_i$ and $\mathcal{N}'_i$. We focus at one agent j, and distinguish between four cases

- $j \notin \mathcal{N}_i \cup \mathcal{N}'_i$: In this trivial case, agent j does not contribute to $f_i(\boldsymbol{x}(t+1)) - f_i(\boldsymbol{x}(t))$, thus i spends 0 for j.
- $j \in \mathcal{N}_i \setminus \mathcal{N}'_i$: j was a neighbor of i at time t but left her neighborhood at time $t+1$. As j can move arbitrarily far, and i does not want to spend much money for agents that are not even in her neighborhood, we bound j's contribution to $f_i(\boldsymbol{x}(t+1))$ by the boundary of i's neighborhood. Therefore, i spends $1 - \|d_{ij}\|_2^2$.
- $j \in \mathcal{N}'_i \setminus \mathcal{N}_i$: j became a neighbor of i at time $t+1$ but was not in her neighborhood at time t. Again, as j could have been arbitrarily far at time t, we bound j's contribution to $f_i(\boldsymbol{x}(t))$ by the boundary of i's neighborhood. Therefore, i spends $\|d'_{ij}\|_2^2 - 1$.
- $j \in \mathcal{N}_i \cap \mathcal{N}'_i$: j was a neighbor of i both at t and $t+1$. In this case, i spends $\|d'_{ij}\|_2^2 - \|d_{ij}\|_2^2$.

Because our protocol forces i to decide how much to spend for all agents at time t, we would like to rework our second case into something similar to the fourth one, to gain a unified spending rule for all agents of i at time t. We then observe that

$$1 - \|d_{ij}\|_2^2 = \|d'_{ij}\|_2^2 - \|d_{ij}\|_2^2 + 1 - \|d'_{ij}\|_2^2 = \|d'_{ij}\|_2^2 - \|d_{ij}\|_2^2 - \left|\|d'_{ij}\|_2^2 - 1\right|$$

which also leads us to transform our third case to

$$\|d'_{ij}\|_2^2 - 1 = -\left|\|d'_{ij}\|_2^2 - 1\right|$$

and, surprisingly, we get a unified rule for all agents leaving or joining i's neighborhood at time $t+1$! Finally, we want to analyze the fourth case as well. We define $\Delta_i = x_i(t+1) - x_i(t)$ and $\Delta_j = x_j(t+1) - x_j(t)$. Since $d'_{ij} = d_{ij} + \Delta_i - \Delta_j$, we get

$$\begin{aligned}\|d'_{ij}\|_2^2 - \|d_{ij}\|_2^2 &= \|d_{ij}\|_2^2 + \|\Delta_i - \Delta_j\|_2^2 + 2d_{ij}^T(\Delta_i - \Delta_j) - \|d_{ij}\|_2^2 \\ &= -2d_{ij}^T\Delta_j + 2d_{ij}^T\Delta_j + \|\Delta_i - \Delta_j\|_2^2 \\ &= -2d_{ij}^T\Delta_j + 2d_{ji}^T\Delta_j + \|\Delta_i - \Delta_j\|_2^2 - 4d_{ij}^T\Delta_i \end{aligned} \tag{5.12}$$

Here, recall the protocol defined at Section 3.2.2, and note that $\left|\|d'_{ij}\|_2^2 - 1\right|$ is exactly the amount of money that i spends (hence the minus sign) if j leaves or joins i's neighborhood at time $t+1$, $-2d_{ij}\Delta_j^T$ is a portion of the amount of money i gives to each agent $j \in \mathcal{N}_i$ and $+2d_{ji}\Delta_j^T$ the equivalent portion of the amount i gets from every $j \in \mathcal{N}_i$, at any time t.

We are now left with only two terms, $\|\Delta_i - \Delta_j\|_2^2 - 4d_{ij}\Delta_i^T$. Recall our relaxation on $K_i(2)$ at (5.5), where we bounded all inertias by their maximum value of 1. However, for agents to have a positive amount of money remaining at time $t+1$, the inertias, along with the cardinality of i's neighborhood, must be included in the model. We recall (3.13) which gives us $\sum_{j\in\mathcal{N}_i} d_{ij} = -\lambda_i^{-1}|\mathcal{N}_i|\Delta_i$, where by λ_i we denote $\lambda_i(t)$. Thus, by summing over all $j \in \mathcal{N}_i$, we can reform our leftover terms to

$$\sum_{j\in\mathcal{N}_i} \|\Delta_i - \Delta_j\|_2^2 + 4\lambda_i^{-1}|\mathcal{N}_i|\|\Delta_i\|_2^2 \tag{5.13}$$

At this point, our analysis is almost over. However, we are not quite there yet. For our argument to hold, we need to relate the amount of money that i spends at time t to a consumption of kinetic 2-energy, and we observe that the term $\sum_{j\in\mathcal{N}_i} \|\Delta_i - \Delta_j\|_2^2$ is very close to the kinetic 2-energy, as defined in Definition 4.31. Thus, for the last time, we rework the term into a consumption of kinetic 2-energy

$$\begin{aligned}\sum_{j\in\mathcal{N}_i} \|\Delta_i - \Delta_j\|_2^2 &= \sum_{j\in\mathcal{N}_i} \|\Delta_i\|_2^2 - 2\Delta_i^T\Delta_j + \|\Delta_j\|_2^2 \\ &= \sum_{j\in\mathcal{N}_i} -\|\Delta_i\|_2^2 - 2\Delta_i^T\Delta_j - \|\Delta_j\|_2^2 + 2\|\Delta_i\|_2^2 + 2\|\Delta_j\|_2^2 \\ &= \sum_{j\in\mathcal{N}_i} -\|\Delta_i + \Delta_j\|_2^2 + 2\|\Delta_i\|_2^2 + 2\|\Delta_j\|_2^2 \end{aligned} \tag{5.14}$$

Now, we observe that we can incorporate the two new terms $2\|\Delta_i\|_2^2 + 2\|\Delta_j\|_2^2$ into the exchange between i and j, if we manage to reverse either term's sign. Since we already have a term relating to the kinetic 2-energy, equal to the amount of money that i has left at time $t+1$, we decide to reverse the first term's sign, and get

$$\begin{aligned}\sum_{j\in\mathcal{N}_i} 2\|\Delta_i\|_2^2 + 2\|\Delta_j\|_2^2 &= \sum_{j\in\mathcal{N}_i}\Big\{2\|\Delta_j\|_2^2 - 2\|\Delta_i\|_2^2 + 4\|\Delta_i\|_2^2\Big\} \\ &= \sum_{j\in\mathcal{N}_i}\Big\{2\|\Delta_j\|_2^2 - 2\|\Delta_i\|_2^2\Big\} + 4|\mathcal{N}_i|\|\Delta_i\|_2^2 \end{aligned} \tag{5.15}$$

We summarize our message-passing protocol, using all of the above equations. At any time step $t \geq 0$, we apply the following two rules to every agent i

- Agent i spends $\|\Delta_i + \Delta_j\|_2^2$ units of money for every $j \in \mathcal{N}_i(t)$, at each time t, and gives to agent j an amount equal to $2(d_{ij} - \Delta_j)^T\Delta_j$.
- For every agent j that becomes, or ceases to be, a neighbor of i at time $t+1$, agent i spends $\Big|\|d'_{ij}\|_2^2 - 1\Big|$.

Finally, we have to provide with an initial "guess" for $f_i(\boldsymbol{x}(0))$. Altough we have that $f_i(\boldsymbol{x}(t)) = \sum_{j\in\mathcal{N}_i(t)}\|d_{ij}\|_2^2$, we want to include all agents in our definition of $f(\boldsymbol{x}(0))$, and provide agent i with some extra money. However, we want to bound the influence of the agents that will not be in $\mathcal{N}_i(0)$, again by the boundary of i's neighborhood. Thus, we arrive at the following initial guess

$$f_i(\boldsymbol{x}(0)) = \sum_{j=1}^{n}\min\Big\{\|x_i(0) - x_j(0)\|_2^2, 1\Big\} \tag{5.16}$$

Now recall the protocol used in Section 3.2.2, and notice that it is exactly the same as the above. We conclude that we have successfully analyzed this protocol in reverse, starting with a simple observation as to what constitutes a bound on the kinetic 2-energy, and explaining our thought process and reasoning at each step. We hope this analysis complements the proof provided in Section 3.2.2 and assists the reader in developing deeper intuition about the model.

5.3 Future Work

In the final section of this thesis, we wish to present several open questions to the reader, that we believe to be significant, thus the target for future work in the field of

opinion dynamics. First of all, while the kinetic 2-energy of the Inertial-HK model is bounded, a strong indication for convergence, it is still not known whether this model converges or not. Furthermore, almost nothing is known about the convergence properties of the Generalized Asymmetric model, and the question relating to the convergence of this model remains one of the most intriguing and important open problems in the field.

The above open problems point to a need for developing clear and practical conditions that distinguish convergent from non-convergent models. Perhaps the most important open question of all is to determine what qualities of the interaction between agents designate a convergent system, and why the lack of one leads to a non-convergent system. Moreover, while convergence has been proven for the HK model, and several of its variants, we do not have any tight bounds on the convergence time of any variant or even the original HK model itself, as of yet.

Besides analyzing and studying existing models, a significant amount of work should be pointed towards developing new, perhaps more complex, models, which provide a better grasp on the real-world interactions between agents and simulate real social networks in a better way. To aid with the analysis of former and new models, novel ideas and techniques also have to be developed. We already demonstrated a novel approach to the analysis of a system in Section 4.5, the concept of a system's s-energy, which raises the question whether more innovative tools like this one could be developed, and lead to better understanding of several model's convergence properties. Finally, we state our belief that many of these questions can be answered with the study of the newly introduced field of influence systems, which generalize the concept of an opinion formation model [14, 15].

Bibliography

[1] Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, Princeton, NJ, USA, 2008. ISBN 0691134405, 9780691134406.

[2] J. R. P. Jr. French. A formal theory of social power. *The Psychological Review*, 63 (3):181–194, 1956. doi: 10.1037/h0046123.

[3] Morris H. DeGroot. Reaching a Consensus. *American Statistical Association*, 69 (345):118–121, 1974. doi: 10.2307/2285509. URL http://www.jstor.org/stable/2285509.

[4] N. E. Friedkin and E. C. Johnsen. Social influence and opinions. In *Mathematical Sociology*, volume 15, pages 193–205, 1990.

[5] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence: models, analysis and simulation. *Artificial Societies and Social Simulation*, 5(3), 2002.

[6] Gérard Weisbuch, Guillaume Deffuant, Frédéric Amblard, and Jean-Pierre Nadal. Meet, discuss, and segregate! *Complexity*, 7(3):55–63, 2002. ISSN 1099-0526. doi: 10.1002/cplx.10031.

[7] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 03(01n04):87–98, 2000. doi: 10.1142/S0219525900000078. URL http://www.worldscientific.com/doi/abs/10.1142/S0219525900000078.

[8] Petter Holme and M. E. J. Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys. Rev. E*, 74:056108, 2006. doi: 10.1103/PhysRevE.74.056108.

[9] Diodato Ferraioli, Paul W. Goldberg, and Carmine Ventre. Decentralized Dynamics for Finite Opinion Games. *Theoretical Computer Science*, 648(C):96–115, 2016. ISSN 0304-3975. doi: 10.1016/j.tcs.2016.08.011.

[10] Ercan Yildiz, Asuman Ozdaglar, Daron Acemoglu, Amin Saberi, and Anna Scaglione. Binary Opinion Dynamics with Stubborn Agents. *ACM Transactions*

on Economics and Computation (TEAC), 1(4):19:1–19:30, 2013. ISSN 2167-8375. doi: 10.1145/2538508.

[11] David Bindel, Jon Kleinberg, and Sigal Oren. How bad is forming your own opinion? *Games and Economic Behavior*, 92(C):248–265, 2015.

[12] Timur Kuran and William H. Sandholm. Cultural Integration and Its Discontents. *The Review of Economic Studies*, 75(1):201–228, 2008.

[13] J. Hagenbach and F. Koessler. Strategic Communication Networks. *The Review of Economic Studies*, 77(3):1072–1099, 2010.

[14] Bernard Chazelle. The Dynamics of Influence Systems. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS '12*, pages 311–320, 2012. doi: 10.1109/FOCS.2012.70.

[15] Bernard Chazelle. Diffusive Influence Systems. *SIAM J. Comput.*, 44(5):1403–1442, 2015. doi: 10.1137/120882640.

[16] Bernard Chazelle. The Challenges of Natural Algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, 2016. ISBN 978-1-4503-4206-3. doi: 10.1145/2908812.2908959.

[17] Bernard Chazelle. Natural Algorithms and Influence Systems. *Communications of the ACM*, 55(12):101–110, 2012. ISSN 0001-0782. doi: 10.1145/2380656.2380679.

[18] Arthur Carvalho and Kate Larson. A Consensual Linear Opinion Pool. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, pages 2518–2524, 2013. ISBN 978-1-57735-633-2.

[19] Alan Tsang and Kate Larson. Opinion Dynamics of Skeptical Agents. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '14, pages 277–284, 2014. ISBN 978-1-4503-2738-1.

[20] John Nash. Non-Cooperative Games. *Annals of Mathematics*, 54(2):286–295, 1951. ISSN 0003486X. doi: 10.2307/1969529. URL http://www.jstor.org/stable/1969529.

[21] Gilbert Strang. *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA, 1993.

[22] Javad Ghaderi and R. Srikant. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(12):3209–3215, 2014. doi: 10.1016/j.automatica.2014.10.034. URL http://dx.doi.org/10.1016/j.automatica.2014.10.034.

[23] Vincent D. Blondel, Julien M. Hendrickx, and John N. Tsitsiklis. On Krause's multi-agent consensus model with state-dependent connectivity. *IEEE Transactions on Automatic Control*, 54(11):2586–2597, 2009. URL http://dblp.uni-trier.de/db/journals/tac/tac54.html#BlondelHT09.

[24] Jan Lorenz. A stabilization theorem for dynamics of continuous opinions. *Physica A: Statistical Mechanics and its Applications*, 355(1):217 – 223, 2005. ISSN 0378-4371. doi: http://dx.doi.org/10.1016/j.physa.2005.02.086.

[25] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005. ISSN 0018-9286.

[26] Arnab Bhattacharyya, Mark Braverman, Bernard Chazelle, and Huy L. Nguyen. On the convergence of the Hegselmann-Krause system. In *Innovations in Theoretical Computer Science, ITCS '13*, pages 61–66, 2013. doi: 10.1145/2422436.2422446. URL http://doi.acm.org/10.1145/2422436.2422446.

[27] Peter Hegarty, Anders Martinsson, and Edvin Wedin. The Hegselmann-Krause dynamics on the circle converge. *Journal of Difference Equations and Applications*, 22(11):1720–1731, 2016. doi: 10.1080/10236198.2016.1235703.

[28] Jan Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838, 2007. doi: 10.1142/S0129183107011789.

[29] Jiangbo Zhang and Ge Chen. Convergence rate of the asymmetric Deffuant-Weisbuch dynamics. *Journal of Systems Science and Complexity*, 28(4):773–787, 2015. ISSN 1559-7067. doi: 10.1007/s11424-015-3240-z. URL http://dx.doi.org/10.1007/s11424-015-3240-z.

[30] Dimitris Fotakis, Dimitris Palyvos-Giannas, and Stratis Skoulakis. Opinion Dynamics with Local Interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016*, pages 279–285, 2016. URL http://www.ijcai.org/Abstract/16/047.

[31] Bernard Chazelle and Chu Wang. Inertial Hegselmann-Krause Systems. *CoRR*, abs/1502.03332, 2015. URL http://arxiv.org/abs/1502.03332.

[32] Jan Lorenz. Heterogeneous Bounds of Confidence: Meet, Discuss and Find Consensus! *Complexity*, 15(4):43–52, 2010. ISSN 1076-2787. doi: 10.1002/cplx.v15:4.

[33] Bernard Chazelle. The Total s-Energy of a Multiagent System. *SIAM J. Control and Optimization*, 49(4):1680–1706, 2011. doi: 10.1137/100791671.

[34] Kshipra Bhawalkar, Sreenivas Gollapudi, and Kamesh Munagala. Coevolutionary opinion formation games. In *Symposium on Theory of Computing Conference, STOC '13*, pages 41–50, 2013. doi: 10.1145/2488608.2488615. URL http://doi.acm.org/10.1145/2488608.2488615.

[35] J. B. Rosen. Existence and Uniqueness of Equilibrium Points for Concave N-Person Games. *Econometrica*, 33(3):520–534, 1965.

[36] Robert W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973. ISSN 1432-1270. doi: 10.1007/BF01737559.

[37] Dov Monderer and Lloyd S. Shapley. Potential Games. *Games and Economic Behavior*, 14(1):124–143, 1996. ISSN 0899-8256. doi: http://dx.doi.org/10.1006/game.1996.0044.

[38] K. J. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3):265–290, 1954. doi: 10.2307/1907353.

[39] Vasco Brattka, Stéphane Le Roux, and Arno Pauly. On the computational content of the Brouwer fixed point theorem. In *Conference on Computability in Europe*, pages 56–67. Springer, 2012.

[40] Mark Yoseloff. Topologic proofs of some combinatorial theorems. *Journal of Combinatorial Theory, Series A*, 17(1):95 – 111, 1974. doi: http://dx.doi.org/10.1016/0097-3165(74)90031-4.

[41] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The Complexity of Computing a Nash Equilibrium. *SIAM Journal on Computing*, 39 (1):195–259, 2009. doi: 10.1137/070699652.

[42] Shizuo Kakutani. A generalization of Brouwer's fixed point theorem. *Duke Mathematical Journal*, 8(3):457–459, 1941. doi: 10.1215/S0012-7094-41-00838-4.

[43] Eyal Even-dar, Yishay Mansour, and Uri Nadav. On the Convergence of Regret Minimization Dynamics in Concave Games. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 523–532, 2009. ISBN 978-1-60558-506-2. doi: 10.1145/1536414.1536486.

[44] Lloyd S. Shapley. A Solution Containing an Arbitrary Closed Component. *Annals of Mathematical Studies*, 40:87–93, 1959.

[45] W. Karush. Minima of Functions of Several Variables with Inequalities as Side Constraints. Master's thesis, Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois, 1939.

[46] H. W. Kuhn and A. W. Tucker. Nonlinear Programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492. University of California Press, 1951. URL http://projecteuclid.org/euclid.bsmsp/1200500249.

[47] Philip Wolfe. Convergence Conditions for Ascent Methods. *SIAM Review*, 11(2): 226–235, 1969. ISSN 00361445.

[48] Stratis Skoulakis. personal communication, 2016.

[49] J.M. Hendrickx and V.D. Blondel. Convergence of different linear and non-linear Vicsek models. In *Proc. 17th International Symposium on Mathematical Theory of Networks and Systems (MTNS '06), Kyoto (Japan)*, pages 1229–1240, July 2006.