On the Robustness of Different Electoral Systems to External Influences

Givi, Glory Mary

This publication URL: https://archive-ouverte.unige.ch//unige:174467
Publication DOI: 10.13097/archive-ouverte/unige:174467

UNIVERSITÉ DE GENÈVE
Departement de physique

FACULTÉ DES SCIENCES
Professeur P. Jacquod

On the Robustness of Different Electoral Systems to External Influences

THÈSE

Presentée à la Faculté des sciences de l'Université de Genève
Pour obtenir le grade de Docteur ès sciences, mention Physique

Par

Glory Mary GIVI

de

Kochi (INDE)

Thèse N° 5791

GENÈVE

Nom de l'imprimerie

2023

Out of clutter, find Simplicity.
From discord, find Harmony.
In the middle of difficulty, lies
Opportunity.

Albert Einstein

Dedicated to my parents and little brother...!

Acknowledgment

I am deeply grateful to all those who have supported and guided me throughout this incredible journey of pursuing my PhD. This thesis represents the culmination of years of hard work and dedication, and I could not have accomplished it without the help and encouragement of many individuals.

First and foremost, I would also like to extend my heartfelt thanks to the almighty, whose grace and blessings have guided me through the challenges of this academic endeavor, filling me with the strength and perseverance to overcome obstacles.

I would like to express my heartfelt gratitude to my advisor, Prof. Philippe Jacquod, for their unwavering support, invaluable guidance, and patience throughout this research. His expertise, mentorship, and commitment to excellence have been instrumental in shaping the direction of my work and my growth as a researcher. I would like to thank Prof. Dirk van der Marel, for being a great mentor, support and guidance. His advice has helped me improve my work and build confidence. Thanks to Robin for teaching me how to get started in a project, collaborations, advice and suggestions regarding my thesis. Overall being an unofficial co-supervisor and go to person when I'm stuck.

Thanks to my colleagues (and ex-colleagues) Koen, Julian, Marc, Laurent and Melvyn, who have provided a supportive and collaborative atmosphere. A special mention to Julian for proofreading my articles and Laurent for helping me get started in Switzerland. A special thanks to Prof. Sangeetha George and Prof. Sudev N.K for all their guidance during these years at a professional and personal level. Many thanks to my friends here: Anita, Rifat, Sauradeep, Naveen, and Pullah. Your friendship has been an essential part of my life during this period.

To my family, especially my parents, Givi and Bini, I can't thank you enough for your love, support, understanding, and unwavering belief in my abilities. Your sacrifices and encouragement have been my pillars of strength, and moulding me to who what I am today. Many thanks to my little brother, Gody, for all your support and showing me the hopes for tomorrow. Last, but not least, thank you to my partner Jerry for your patience and support through these last months of my thesis; you made a big difference.

Lastly, I express my gratitude to all those unnamed individuals who have, directly or indirectly, contributed to this thesis. Your collective influence on my academic and personal journey is immeasurable. This thesis is a testament to the support and guidance I have received from these wonderful people. While the list of names here may be limited, the impact of their contributions on my life and research is immeasurable.

Thank you all for being a part of this remarkable journey.

Abstract

One of the major endeavors of science and engineering is to study and predict the behavior of large dynamical systems. It includes a myriad of problems–from weather prediction to understanding the dynamics of artificial neural networks. Although several important problems in physics and engineering fall under the ambit of large dynamical systems, carefully designed large dynamical systems provide insights into social phenomena as well. This thesis is concerned with the evolution of opinions among voters during election processes, which comes under the broad umbrella of computational social sciences. We employ dynamical systems and methodologies developed in physics and mathematics to model such election processes.

We investigate the opinion formation process among an ensemble of interacting individuals modeled by a dynamical system. The problem we form on is the following: There have been numerous recent instances of social networks being actively utilized to influence elections by disseminating rumors and misleading information. The possibility of manipulating elections using external influence directly raises the question of the intrinsic robustness of existing electoral systems. Identifying the social characteristics of a voter population that make it resilient to opinion manipulation is particularly interesting. Determining which among the existing democratic electoral systems is more resilient to external influences is equally important. We address these questions using a well-designed opinion dynamics model.

Building on earlier works in computational social sciences, we adapt a mathematical model of opinion dynamics, where agents represent voters interacting with one another. The model includes (i) a *natural opinion* that represents the opinion of each voter in the absence of interaction with other voters, (ii) a *confidence bound* that determines whether a voter interacts with others, (iii) an external influence field trying to change the voters' opinion, and (iv) a differential equation governing the time evolution of the opinion of voters.

There are many models of opinion dynamics, however only a few have been validated with real data. The validation of the models is challenging due to the need for more real data that can be transformed into a metric that can quantitatively or even qualitatively validate these models. One of the main results is the validation of our model. We validate the model we constructed using historical data of the US House of Representative elections from 2012 to 2020. The model captures qualitatively, if not quantitatively, the US House of Representatives election's volatility, which we consider a solid validation of our approach.

We evaluate the robustness of different electoral systems to external influences. An external influence is introduced in the system as an influence field to change the election outcome. We then quantify the total effort the influence field has to exert to change the electoral outcome as a function of agent polarization, confidence bound, different electoral systems, and the number of political parties. We present a detailed study on the robustness of bipartite electoral systems (i.e., with two major parties) and extend it to multipartite (i.e., with multiple major parties) electoral systems.

For a bipartite system, we find that voter populations are more resilient against external opinion manipulations if they have less polarized initial opinion distribution and when agents interact with each other regardless of their opinion differences. We extend the model to a

multipartite system, including the left-right political ordering for the opinion of each agent, and this ordering significantly affects the system's dynamics. For a multipartite system, the voter population's resilience depends on the runner-up party's position in the left-right political spectrum. When agents interact with each other regardless of their opinion differences, the party at the centre wins the elections, and voter populations become more resilient against external opinion manipulations.

Finally, the outcome of an election depends on how the individual votes are aggregated to elect representatives. This aggregation process depends on the electoral system under consideration, such as proportional, single representation, winner-takes-all, and so forth corresponding to different countries and elections for different branches of governance. Applying our model to different electoral systems, we find that the proportional representation are the most robust against external influences.

All our results and conclusions are independent of the number of existing political parties. Our results emphasize the need to encourage public debates and discussions during political campaigns to strengthen democratic processes. However, different electoral processes behave remarkably differently under the influence of external manipulation, and proportional representation makes for a more robust electoral process. This thesis can be extended to include all other different electoral systems, including two-round system, block-voting and so forth.

Résumé

L'un des principaux objectifs de la science et de l'ingénierie est d'étudier et de prédire le comportement de grands systèmes dynamiques. Cette tâche comprend une grande variété de problèmes, allant des prévisions météorologiques à la compréhension de la dynamique des réseaux neuronaux artificiels. Bien que souvent les grands systèmes dynamiques servent à décrire des problèmes importants de la physique et de l'ingénierie, de tels systèmes soigneusement conçus permettent également de mieux comprendre les phénomènes sociaux. Cette thèse porte sur l'évolution des opinions des électeurs au cours des processus électoraux, à travers le prisme des sciences sociales computationnelles (computational social sciences). Nous utilisons des systèmes dynamiques et des méthodologies développées en physique et en mathématiques pour modéliser de tels processus électoraux.

Nous étudions le processus de formation de l'opinion parmi un ensemble d'individus en interaction, modélisés par un système dynamique. Le problème sur lequel nous nous penchons est le suivant : Il y a eu de nombreux cas récents d'utilisation active des réseaux sociaux pour influencer les élections, en diffusant des rumeurs et des fausses informations. La possibilité de manipuler des élections en utilisant une influence extérieure soulève directement la question de la robustesse intrinsèque des systèmes électoraux existants. Si l'identification des caractéristiques sociales d'une population d'électeurs qui la rend résistante à la manipulation de l'opinion est particulièrement intéressante, déterminer lequel des systèmes électoraux démocratiques existants est le plus résistant à la manipulation l'est tout autant. Nous abordons ces questions à l'aide d'un modèle de dynamique de l'opinion approprié.

En nous appuyant sur des travaux antérieurs dans le domaine des sciences sociales computationnelles, nous adaptons un modèle mathématique de la dynamique de l'opinion, où les agents représentent les électeurs interagissant les uns avec les autres. Le modèle comprend (i) une opinion naturelle qui représente l'opinion de chaque électeur en l'absence d'interaction avec d'autres électeurs, (ii) un seuil de confiance qui détermine si un électeur interagit avec les autres, (iii) un champ d'influence externe essayant de changer l'opinion des électeurs, et (iv) des équations différentielles régissant la dynamique de l'opinion de chaque électeur.

Il existe de nombreux modèles de dynamique de l'opinion, mais seuls quelques-uns ont été validés avec des données réelles. La validation des modèles est un défi en raison de la nécessité de disposer d'un grand nombre de données réelles pouvant être transformées en une métrique capable de valider quantitativement ou même qualitativement ces modèles. L'un des principaux résultats de ce travail est la validation de notre modèle. Nous validons le modèle que nous avons construit en utilisant les données historiques des élections à la Chambre des Représentants des États-Unis de 2012 à 2020. Le modèle capture qualitativement, sinon quantitativement, la volatilité de ces élections, ce que nous considérons comme une validation solide de notre approche.

Nous évaluons ensuite la robustesse de différents systèmes électoraux aux influences externes. Une influence externe est introduite dans le système sous la forme d'un champ d'influence afin de modifier le résultat de l'élection. Nous quantifions ensuite l'effort total que le champ d'influence doit déployer pour modifier le résultat électoral en fonction de divers paramètres : la polarisation

des agents, le seuil de confiance, le système électoral et le nombre de partis politiques. Nous présentons une étude détaillée sur la robustesse des systèmes électoraux bipartites (c'est-à-dire avec deux grands partis) et l'étendons aux systèmes électoraux multipartites (c'est-à-dire avec plusieurs grands partis).

Dans le cas d'un système bipartite, nous constatons que les populations d'électeurs sont plus résistantes aux manipulations externes de l'opinion si elles sont moins polarisées et si les agents interagissent les uns avec les autres indépendamment de leurs différences d'opinion. Nous étendons le modèle à un système multipartite, en incluant l'orientation politique gauche-droite pour l'opinion de chaque agent. Cette orientation affecte de manière significative la dynamique du système. Pour un système multipartite, la résilience de la population dépend de la position du parti arrivé en deuxième position dans le spectre politique gauche-droite. Lorsque les agents interagissent les uns avec les autres indépendamment de leurs divergences d'opinion, les partis centristes gagnent majoritairement les élections, et les populations d'électeurs deviennent plus résilientes face aux manipulations de l'opinion.

Enfin, le résultat d'une élection dépend de la manière dont les votes individuels sont agrégés pour élire les représentants. Ce processus d'agrégation dépend du système électoral considéré, tel que le scrutin majoritaire proportionnel, le scrutin majoritaire absolu, le système "winner-takes-all", correspondant à différents pays et à des élections pour différentes branches du gouvernement. En appliquant notre modèle à différents systèmes électoraux, nous constatons que la représentation proportionnelle est la plus robuste face aux influences extérieures.

Tous nos résultats et conclusions sont indépendants du nombre de partis politiques existants. Nos résultats soulignent la nécessité d'encourager les débats publics et les discussions pendant les campagnes politiques afin de renforcer les processus démocratiques. Cependant, les différents processus électoraux se comportent remarquablement différemment sous l'influence de manipulations externes, et le scrutin proportionnel constitue un processus électoral plus robuste. Cette thèse peut être étendue à tous les autres systèmes électoraux, y compris le système à deux tours, le vote en bloc, etc.

Publications

The following publication is prepared from work presented in this thesis:

1. G. M. Givi, R. Delabays, M. Jacquemet, and P. Jacquod, "On the robustness of democratic electoral processes to computational propaganda", Scientific Reports **14** (2024)

Contents

Acknowledgment v

Abstract vii

Résumé ix

Publications xi

1 Opinion Dynamics Models and Their Applications 1
- 1.1 Opinion Dynamics Models . . . 4
  - 1.1.1 French - DeGroot Model . . . 4
  - 1.1.2 Friedkin -Johnson's model . . . 6
  - 1.1.3 Abelson and Taylor model . . . 7
  - 1.1.4 Confidence-Bound Models . . . 8
  - 1.1.5 Voter Model . . . 16
  - 1.1.6 The Majority Rule Model . . . 19
  - 1.1.7 Sznajd Model . . . 22
  - 1.1.8 Multidimensional Models . . . 24
- 1.2 Interaction Networks . . . 28
- 1.3 Validation . . . 32
- 1.4 External Influence . . . 34
- 1.5 Electoral Systems and Computational Propaganda . . . 37
- 1.6 Appendix . . . 42
  - 1.6.1 Linear Algebra Concepts . . . 42

2 Two-party System 43
- 2.1 Opinion Dynamics . . . 43
  - 2.1.1 Natural Opinions . . . 44
  - 2.1.2 External Influence . . . 45
- 2.2 Electoral Outcomes . . . 46
  - 2.2.1 Electoral Units . . . 46
  - 2.2.2 Electoral Systems . . . 47
- 2.3 Validation of the Model . . . 47
- 2.4 Robustness Analysis . . . 51
  - 2.4.1 Robustness Against Polarization and Bias . . . 52
  - 2.4.2 Robustness of Electoral Systems . . . 56
- 2.5 A case study on United States election data . . . 58
- 2.6 Conclusion . . . 59
- 2.7 Appendix . . . 61
  - 2.7.1 Influence strategies on single electoral unit . . . 63
  - 2.7.2 Influence strategies for multiple electoral units . . . 64

3 Multi-party System **67**
3.1 Opinion Dynamics . . . 67
3.1.1 The Opinion Space . . . 68
3.1.2 Natural Opinions . . . 69
3.1.3 External Influence . . . 69
3.2 Result of Elections . . . 71
3.3 Single Electoral Unit . . . 71
3.3.1 Plurality System . . . 72
3.3.2 Ranked Choice Voting System . . . 75
3.4 Multiple Electoral Units . . . 79
3.5 Conclusion . . . 80
3.6 Appendix . . . 82
3.6.1 Equivalence to One-Dimensional Dynamics . . . 82
3.6.2 Effect of Ordering . . . 84
3.6.3 Opinion Volume Balancing . . . 87

4 Conclusion **93**

Bibliography **97**

1 Opinion Dynamics Models and Their Applications

Complex systems are omnipresent in everyday life in the form of power grids, transportation, communication system, and in the society itself. In general, the interaction between an ensemble of individual entities composing the system are captured by complex networks, and is fundamental to the behavior of the system itself. The interactions greatly impact the way entities are functioning and evolving as a function of time and has presented us with fascinating phenomena such as phase transitions. The exploration of complex systems holds significant importance within the realms of both physics and computational social science. The studies on complex systems aim to characterize the macroscopic behavior of the system by assigning simple, intuitive rules to the interaction between an ensemble of entities. In physics the constituents are degrees of freedom of particles, whereas in computational social science it is attributes of individuals in a society.

This thesis studies opinion formation among voters during election processes. The study comes under the scope of computational social sciences. The latter refers to the study of social phenomena using mathematical methods. Examples include computational modeling of economic systems [2], social media content analysis [3], elections [4] etc. It helps us to investigate and predict common behavioral patterns that we observe in society. Studies in computational social sciences employ dynamical systems to explore how interactions lead to behavioral patterns depending on different initial conditions and system parameters.

A dynamical system is a mathematical formalism used to describe the time evolution of systems. In general, it is described by an ordinary differential equations (for continuous time) or iterated maps (also known as difference equations) [5]. The ODE can be formulated as

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, t), \tag{1.1}$$

where $\boldsymbol{x}$ represents the coordinates vector corresponding to the quantity of interest in the system. The derivative $\dot{\boldsymbol{x}}$ represents the rate of change of the state vector $\boldsymbol{x}$ with respect to time t. The function $\boldsymbol{f}$ defines how the variables in $\boldsymbol{x}$ change over time.

In the context of opinion formation during an election process, the quantities of interest are the opinions (represented by a vector of real numbers) of voters. Here, the dynamical system is constituted by a coupled system of ordinary differential equations (ODEs) with the opinions of individuals as the elements evolving in time due to interactions and other external influences like media. The complexity of the dynamical system increases as we strive to capture more details in the model. This leads to different kinds of dynamical systems such as autonomous or non-autonomous, linear or nonlinear, etc. If the ODEs does not explicitly depend on the time, then it is called an autonomous system; otherwise, non-autonomous. If the dynamical system is represented by a linear ODE, then we call the system to be linear; otherwise nonlinear. We form an autonomous, linear model for opinion formation during election processes.

Intuitive ideas about opinion formation can be conveniently translated into mathematical equations that govern the rate of change of opinions. This makes the dynamical systems approach suitable for studying opinion formation during election processes. The dynamical system approach provides us with the complete time evolution of the quantities of interest.

An individual's opinion formation depends on many factors. The opinion formation can be the result of *intrinsic* or *extrinsic* experiences. Intrinsic experiences refer to personal experiences, upbringing, education, and so on. Whereas extrinsic experiences refer to the interaction with other individuals in society, information exchanges via social media, newspapers, etc., or even a targeted effort to manipulate the opinion by other individuals or social medias. The ability of individuals to adapt and change in light of new experiences enables lifelong learning and adaption of their opinions. The evolution of opinions of different individuals subjected to the same experiences may not be the same. This uncertainty in opinion change of individuals results from how each individual comprehends information and experiences.

It is evident that the opinion formation of an individual is a complex process, and understanding them at the individual level is impossible using mathematical models. Even if we succeed in creating a perfect model for an individual's opinion formation, simulating the collective opinion over a large population would be computationally demanding.

A more feasible approach is to study the collective opinions of a population using statistics, which can help us understand opinion formation qualitatively. Statistics has been successful in modeling physical systems, and its general conceptual framework has been applied to fields like biology, medicine, information technology, and computer science. Statistics can provide insights into the dynamics of social systems and their behavior at a macroscopic level [6]. This is particularly relevant while studying election processes as the election outcome is determined by an aggregate opinion of individuals. This is the approach we follow.

In societies, the basic constituents are individuals and they exhibit interesting behaviors due to their interactions [7]. Developing a quantitative understanding of how these behaviors in opinions arise among individuals have always been of interest to scientists and philosophers [8]. For example, even though the individuals have very different opinions before interactions, the opinions of individuals converge towards one another after interactions. Understanding the emergence of these behaviors is an important objective of the statistical approach. Statistics capture the emergence of these phenomena by considering ensembles of systems with initial conditions drawn from a probability distribution.

Dynamical models that are specifically used to study the evolution of opinions and their regularities at a large scale are called *opinion dynamics models*. In general, opinion dynamics models can be described as a system that involves a set of agents, a network representing their interactions, any external influences, and an equation that governs the evolution of an agent's opinion as a function of time. This is concisely shown in Fig. 1.1, taken from Ref. [9] (interactions are called as *fusion rules* in this paper). Mathematical approaches to opinion evolution date back to 1956 to the work of French [10]. DeGroot complemented this work [11], yielding the well-established French-DeGroot model, where agents update their opinions at discrete time steps by the weighted average of their neighbors' opinions.

Another early work on opinion dynamics was proposed by Abelson [12] in 1964, which is a continuous-time version of the French-DeGroot model. Taylor [13] extended this model in 1968 to include constant communication sources like mass media, non-linear models, and variable interaction rates. In a similar spirit, Friedkin and Johnson defined their model in 1990 [14], by adding the effect of prior belief into the French-DeGroot model. A detailed description of these models is given in the subsequent sections. Since then, many models have been proposed with various opinion formats and interaction rules among the agents to describe experimentally observed opinion patterns, such as consensus, polarization, or fragmentation [15]. In the case

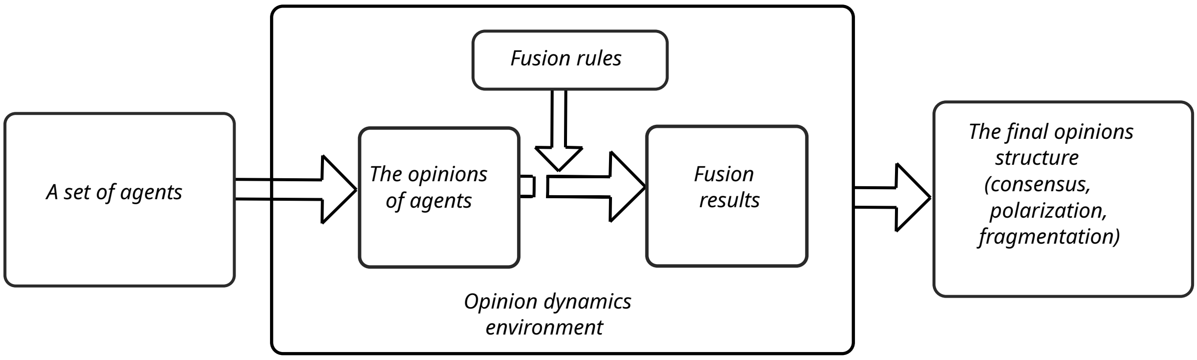


Figure 1.1: The framework of fusion process in opinion dynamics. *Figure taken from Ref. [9]*

of consensus, all individuals align to a single opinion, whereas for polarization there are two cluster of individuals with different opinions, and for fragmentation, there are multiple clusters of individuals with different opinions. These models allow us to extract trends or phenomena and generally not to predict any specific event such as election outcome. The phenomena of interest can vary from the emergence of fads, collective decision-making, reaching consensus, polarization or fragmentation, minority opinion spreading, the role of zealots, minority opinion survival, electoral processes and so on [16].

The main focus of this thesis is the application of opinion dynamics models to electoral processes. The outcome of an election is a result of aggregation of votes, which comes under one of the branches of statistics. One common type of opinion dynamics model used in the study of elections is the voter model [17] (The voter model will be discussed in detail later in this chapter). In the voter model, each agent has a initial opinion, a susceptibility to external influence, and interactions with other voters. These models can capture the effects of various factors such as media influence, targeted advertising, social networks, and ideological polarization, on the evolution of opinions within a population [18, 19].

Recently, election manipulations have become increasingly frequent [20], necessitating considerable attention from scholars and policymakers alike. The very foundation of democratic governance is at stake if the integrity of elections is compromised. It is imperative to recognize the importance of this issue and delve into its complexities to devise effective safeguards against manipulation. Understanding the multifaceted nature of election manipulations which encompasses technological advancements and psychological influences, is paramount. With technological advancements, social media is one of the main drivers of spreading misinformation.

In modern elections, social networks play a crucial role in forming public opinion [21]. They help politicians reach out to people, get support for causes, and quickly share messages over a wide area. Social media platforms allow disinformation tactics in addition to disseminating political ideas. While some of these disinformation campaigns are run directly by individuals, most are run by computer programs known as bots [22, 23]. Bots are designed to be adept at manipulating trends and to flood networks with spam. Controlled and digitally organized fake news may modify voters' opinions and how they vote. In extreme cases, this can change the outcome of an election. Indeed, there are several recent, well-identified cases of social networks being intensively used to influence electoral processes by spreading propaganda and false information.

This malicious influence using social networks directly raises the question of the intrinsic robustness of electoral processes against external attempts to manipulate voter opinions.

Particularly interesting is to identify the social characteristics of a voter population that renders it more resilient against opinion manipulation. Equally important is to determine which of the existing democratic electoral systems is more robust to external influences. In this thesis we construct a novel opinion dynamics model to address these questions.

Before we discuss our model and results, the rest of this chapter is devoted to the state-of-the-art and the discussion of existing models of opinion dynamics such as *French-DeGroot model*, *Friedkin-Johnson model*, *Abelson and Taylor model*, *Confidence-Bound models*, *Voter model*, *Sznajd model*, *Majority Rule model*, and so on.

This chapter is organized as follows. In Sec. 1.1, we present different opinion dynamics models and their extensions. In Sec. 1.2, we provide an introduction to the fundamentals and construction of interaction networks. Sec. 1.3 presents the existing literature on validating the models using some historical data. Sec. 1.4 discusses the impact of different types of external influences on opinion dynamics models. Sec. 1.5 offers a comprehensive review of computational propaganda, different types of democratic electoral systems and gaps in the existing research. Sec. 1.6 discusses the linear algebra concepts required for understanding the literature and key theories in the existing research.

1.1 Opinion Dynamics Models

Opinion dynamics models are dynamical systems whose degrees of freedom represents agent's opinions. Mathematically, an agent's opinion is defined as a variable that can take values in a set of real numbers. The collective opinion of the population is represented using a vector with individual's opinions as its components. An agent can have an opinion on several different topics. In that case, the opinion of an individual is represented as a vector, with each of its components denoting the individual's opinion on different topics, and an opinion matrix defines the collective opinion of the population. In short, the agent in the network is endowed with an opinion that is a scalar or a vector, which can take values in a discrete set or a continuous interval.

Agents interact with one another and these interactions define an interaction network. We provide a detailed description of the fundamentals of networks and construction of different type of networks in Sec. 1.2. In this section, we summarize some of the main opinion dynamics models and their main results. These models can be extended to study specific problems.

1.1.1 French - DeGroot Model

One of the first models of opinion dynamics was proposed by social psychologist French in 1956 [10]. French considered a small network of four agents with different types of connectivity (disconnected, weakly connected, unilaterally connected, strongly connected, and complete networks), to analyze the influence of agents on each other. They observed a tendency for agents' opinions to converge towards one another, and this effect strongly depends on the connectivity of the network.

In general, the model describes a system in which a group of agents with initial opinions interact with each other to update their opinions over discrete time steps, based on the average opinion of their neighbors. Let there be N agents in the interaction network and the opinion of an agent i at time t be denoted as $x_i(t)$. The key object of the model is a row stochastic (i.e., $W_{ij} \geq 0$ and $\sum_j W_{ij} = 1 \, \forall i$) $N \times N$ influence matrix W. The dynamics of the system is

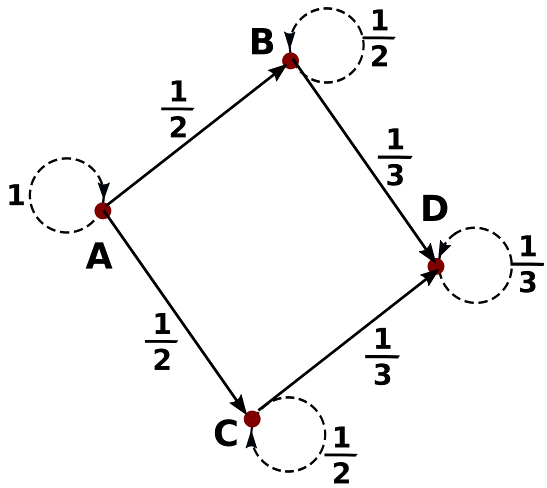


Figure 1.2: An example of French model with $N = 4$ agents

defined as

$$x_i(t+1) = \sum_{j=1}^{N} W_{ij} x_j(t), \tag{1.2}$$

where W_{ij} represents the weight of agent j's opinion at each step of the opinion iteration to agent i's opinion at the next step. In general, the system exhibits an *attractive dynamics*, the tendency of agents' opinions to converge or move closer to each other over time due to the weighted averaging of their neighbors' opinions.

Weight W_{ii} denotes the self-influence weight, i.e., the confidence of the agent or agent's openness to consider other agents' opinions. If $W_{ii} = 0$, it implies that the agents do not have any confidence in their opinion and fully depend on other agents' opinions. However, $W_{ii} = 1$ implies that the agent is very sure of their opinions and does not consider the opinions of their neighbors; thus, $x_i(t) = x_i(0)\ \forall t$. These agents are called *stubborn agents*, *inflexible agents* or *zealots* in the opinion dynamics literature.

French introduced a specific instance of this model. He proposed a simple directed graph G in which the nodes correspond to the agents; each agent is assumed to have some self-confidence and thus have self-loops (see Fig. 1.2). An agent j's opinion influences agent i, or j 'has power over' i, if there is an edge (j, i). At each time step, an agent updates its opinion to the mean value of the opinions of it's neighbors including their own opinion. The weighted graph is shown in Fig. 1.2 and the dynamics corresponds to,

$$\begin{bmatrix} A(t+1) \\ B(t+1) \\ C(t+1) \\ D(t+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} A(t) \\ B(t) \\ C(t) \\ D(t) \end{bmatrix}. \tag{1.3}$$

The ability of an agent to control the behavior of a group is denoted as the *social power* of the agent. French's work demonstrates the influence of social power on opinion formation. Note that agent A is a zealot, and its opinion remains unchanged through the dynamics. Thus agent A increases his influence at the expense of the other agents, and he is considered as the agent with highest social power in graph G.

The agents are said to reach consensus if and only if $\lim_{t\to\infty} x_i(t) = x^* \; \forall i$. If there are two or more different opinions at the final state, the system is said to reach polarization and fragmentation, respectively. The primary goal of French was not to study the dynamics by which the system reaches consensus, polarization, or fragmentation, but instead derive a mathematical model for social power. French proposed various conditions on the connectivity of the interaction networks for reaching a consensus, which were later revised and rigorously proved by Harary using the theory of higher transition probabilities in Markov chains [24].

The model was then generalized by American statistician DeGroot [11]. He described a simple interaction rules called 'iterative opinion pooling' in which agents try to consensus by interaction. The resulting French-DeGroot model is considered the *classical model* of opinion dynamics. If the system reaches consensus, then

$$\lim_{t\to\infty} \boldsymbol{x}(t) = \lim_{t\to\infty} W^t \boldsymbol{x}(0) = x^* \mathbf{1}, \quad x^* \in \mathbb{R}, \tag{1.4}$$

where $\mathbf{1}$ is a vector of all ones with size N.Then it follows from Eq. (1.4) that a consensus is reached if and only if there exist a vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_N)$ such that, for $i = \{1, \cdots, N\}$ and $j = \{1, \cdots, N\}$,

$$\lim_{t\to\infty} W_{ij}^t = \pi_j. \tag{1.5}$$

If Eq. (1.5) is satisfied, then $\pi_1, \cdots, \pi_N$ are non-negative, and $\sum_i \pi_i = 1$, for all i and j. Suppose that a consensus is reached, then we have $\boldsymbol{\pi} W = \boldsymbol{\pi}$, thus the vector $\boldsymbol{\pi}$ is the eigenvector of W corresponding to eigenvalue 1.

Berger [25] presented a necessary and sufficient condition for reaching consensus in the DeGroot model, which states that all agents will form a consensus if and only if there exists $t \in \{1, 2, \cdots\}$ such that the matrix power W^t contains at least one strictly positive column.

1.1.2 Friedkin -Johnson's model

Friedkin and Johnson [14] extended the French-DeGroot model to include *prejudices* in agents' opinion. Here prejudices refer to the prior belief that an agent has. The Friedkin-Johnson (FJ) model computes an agent's opinion as a convex combination of contribution from its neighbors and a prejudice value. The contribution from neighbors is same as that in French-DeGroot model. The update equation for the i^{th} agent is given by

$$x_i(t+1) = \lambda_i \sum_{j=1}^{N} W_{ij} x_j(t) + (1 - \lambda_i) v_i, \quad t = 0, 1, \cdots, \tag{1.6}$$

where $\lambda_i \in [0, 1]$, v_i is the prejudice value, and $x_i(0) = v_i$.

Let $\mathrm{diag}(x)$ denote the diagonal matrix with elements of vector x as its diagonal and let $\Lambda = \mathrm{diag}([\lambda_1, \lambda_2, \cdots, \lambda_N]^T)$. For $\Lambda = \mathbb{I}$, all agents are prejudice-free and Friedkin-Johnson model reduces to French-DeGroot model as in Eq. (1.2). On the other hand, when $\Lambda = 0$, all agents are completely prejudiced and $x_i(t) = v_i$ for all agents i. It is proved in Ref. [26] that the FJ model converges if and only if $\lim_{t\to\infty} (\Lambda W)^t$ exists and the limit $\lim_{t\to\infty} \boldsymbol{x}(t) = \boldsymbol{x}_\infty$ depends on the initial condition $\boldsymbol{x}(0) = \boldsymbol{v}$.

Parsegov [27] analyzed the stability and convergence properties of the Friedkin-Johnson model by dividing the agents into two groups. The first group includes the prejudiced agents

(agents with $\lambda_i \neq 1$) and all the agents influenced by them. All the remaining agents belong to the second group. It is shown that the dynamics of the first group is always convergent and asymptotically stable. The dynamics of the second group is the same as that in the French-DeGroot model as the agents in this group are oblivious to prejudice. In general, the final opinions in FJ model are dependent on prejudices and they seldom reach consensus.

1.1.3 Abelson and Taylor model

In this section, we discuss Abelson's model [12] and its extended version by Taylor [13]. Abelson's model can be considered the continuous time version of the French-DeGroot model. By recalling that $1 - w_{ii} = \sum_{j \neq i} w_{ij}$, the French-Degroot model can be reformulated as

$$x_i(t+1) - x_i(t) = \sum_{j \neq i} w_{ij}[x_j(t) - x_i(t)] \quad \forall i. \tag{1.7}$$

In the continuous-time limit, we obtain Abelson's model [12]

$$\dot{x_i}(t) = \sum_{j \neq i} a_{ij}(x_j(t) - x_i(t)), \quad i = 1, \cdots, n, \tag{1.8}$$

where $A = (a_{ij})$ is a non-negative matrix of influence weight (not necessarily stochastic). The opinion of an agent shifts infinitesimally towards their neighbors' opinions by a weighted superposition of the difference between their own opinion and those of their neighbors. A zealot agent say i corresponds to $a_{ij} = 0$ for all $j \neq i$ and hence $\dot{x_i} = 0$.

The linear Abelson model can be compactly written as

$$\dot{\boldsymbol{x}}(t) = -L\boldsymbol{x}(t), \tag{1.9}$$

where L is the Laplacian matrix of the interaction network ($L_{ij} = -a_{ij}$, for $i \neq j$ and $L_{ii} = \sum_k a_{ik}$). Since the Laplacian is a positive semi-definite matrix, the system always converges. As the Laplacian of the matrix has at least one eigenvalue as zero, the stationary state of the system depends on the initial condition.

Subsequently, Taylor [13] proposed different models to show the effects of external influences such as mass media. Let the influence from m external sources be denoted as $s_1, \cdots, s_m \in \mathbb{R}$. Then the dynamics evolves as

$$\dot{x_i}(t) = \sum_{j=1}^{N} a_{ij}(x_j(t) - x_i(t)) + \sum_{k=1}^{m} b_{ik}(s_k - x_i(t)), \tag{1.10}$$

where b_{ik} is non-negative. It is not necessary that all the agents should be influenced by the external sources, (i.e. $b_{i1} = \cdots = b_{im} = 0$), but agents with $\sum_{k=1}^{m} b_{ik} > 0$ are influenced by one or more communication sources. Taylor has demonstrated that the presence of communication sources induces opinion polarization. Here equilibrium is defined as the value of vector $\boldsymbol{x}$ at $\dot{\boldsymbol{x}} = 0$. In general, the system converges to the unique equilibrium, characterized by $s_1, \cdots, s_k$.

Some other extensions in Taylor's paper [13] include the introduction of variable resistance (i.e., the more extreme the opinion of an agent is, the more difficult it is to convince them) and variable interaction rate (i.e., the greater the difference between their opinion is, weaker the interaction between them) and its stability conditions.

A variant of Taylor models is used to study the effect of zealots on the overall opinion dynamics of networked systems [28]. Zealots with prejudiced opinions towards a particular opinion significantly impact the collective opinion formation process. The authors generate mathematical quantities representing the system's transient dynamics and emergent stationary opinion states using a modified version of the network Laplacian. The study provides a characterization of the ability of a single zealot to change the prevailing opinion coherently. The study also includes the opinion heterogeneity of the emergent non-consensus states for two zealots and describes their statistical features using a graph metric akin to the resistance distance in electrical networks.

1.1.4 Confidence-Bound Models

The proverb 'birds of a feather flock together' illustrates the 'homophily principle'. It states that agents with similar opinions interact more frequently and intensely with one another than ones that are different. The homophily principle is mathematically captured in confidence-bound models using a real number that defines the interactions between the agents depending on their similarity in opinions.

Consider a population of agents with different individual opinions about an issue. Agents in the model are influenced by other agents in the population only if the difference between their opinions is below a certain value ϵ. Here, the value ϵ is called the confidence bound, also known as 'communication distance', 'confidence threshold', or 'confidence strength' in the literature. Thus agents interact with other agents only if their opinions are similar.

Confidence-bound models were first investigated independently by Deffuant and Weisbuch [29], and Hegselmann and Krause [15], which will be discussed later in this section. The Deffuant - Weisbuch (DW) and Hegselmann-Krause (HK) models rely on repeated averaging under the confidence bound but differ in their interaction regimes. The DW model is better suited for modeling one-to-one interaction in large populations, and the HK models are suited for modeling public meetings, where interaction occurs in a group. If the confidence bound is the same for all the agents in the population, then the model is called homogeneous; otherwise, it is heterogeneous [30].

Deffuant-Weisbuch (DW) Model

In the DW model, the opinion of each agent in the interaction network lies in the continuous interval $[0, 1]$, where 0 denotes complete disagreement, and 1 denotes complete agreement to the issue. At each time step, two agents i and j are selected randomly. The agents i and j interact with each other only if the difference between their opinions is smaller than the confidence bound ϵ, i.e. if $|x_i(t) - x_j(t)| \leq \epsilon$. Otherwise, another pair of agents is selected. If they interact, they adjust their opinions according to

$$\begin{cases} x_i(t+1) = x_i(t) + \mu(x_j(t) - x_i(t)) \\ x_j(t+1) = x_j(t) + \mu(x_i(t) - x_j(t)) \end{cases} , \; t = 0, 1, \cdots \tag{1.11}$$

where $\mu \in [0, 1]$ is the known as the convergence parameter. The parameter μ controls the convergence rate of agent i towards agent j. The average opinion of any pair of agents in the population is the same before and after their interactions. Thus, the population's global average remains invariant under DW model. If $\mu = 0$, then two agents will not change their opinions

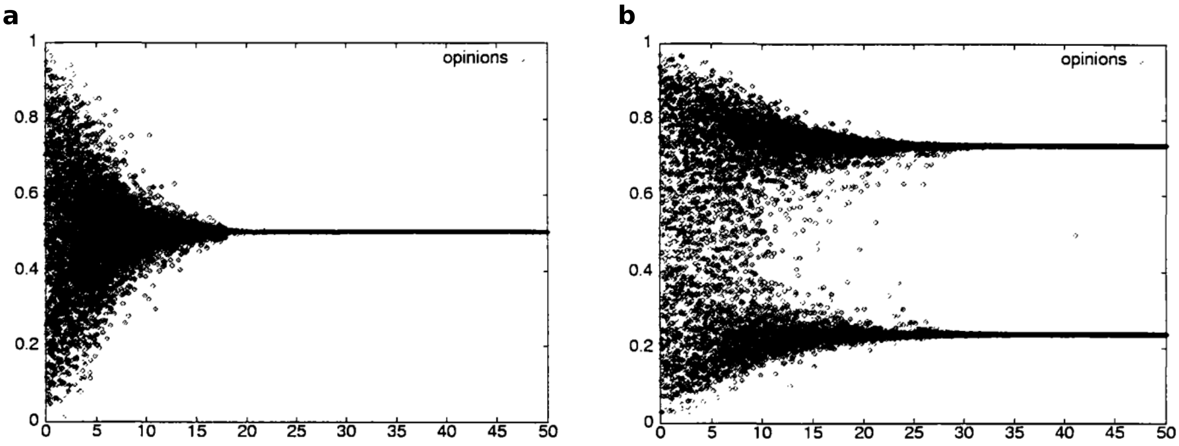


Figure 1.3: Time chart of opinions. One time unit corresponds to sampling 1000 pair agents. **a**: $[\epsilon = 0.5, \mu = 0.5, N = 2000]$, **b**: $[\epsilon = 0.2, \mu = 0.5, N = 1000]$. *Figure and caption taken from [29]*

after interactions, whereas if $\mu = 1$, the agents exchange their opinions and with $\mu = 0.5$, the agents will reach the average of their opinion at the next time step.

The opinions of agents continue to evolve until agents reach their final opinions with one or more clusters (well-separated region in the final opinions). The agents from different clusters have a difference in opinions greater than the confidence bound thereby preventing interactions between them. The time evolution of the DW model in the original work of Deffuant and Weisbuch [29] is shown in Fig. 1.3. Initial opinions are randomly generated from the uniform distribution on $[0, 1]$. Each time unit corresponds to a sampling of 1000 pairs of agents.

The system evolve according to Eq. (1.11) and the opinions of agents converge, but consensus is only achieved for larger values of ϵ. For smaller values of ϵ, the system evolves to form polarization with two stable opinions (refer Fig. 1.3 **b**) or fragmentation with many stable opinions. Qualitatively, the dynamics mainly depend on the confidence bound ϵ.

Monte Carlo simulation reveals that the opinion distribution of agents converge to one or more clusters ($c \approx \lfloor \frac{1}{2\epsilon} \rfloor$, where c is the number of clusters) depending on the value of confidence bound ϵ. As a result, there can be no other cluster within a 2ϵ interval centered on a cluster. Thus the ratio $\lfloor \frac{1}{2\epsilon} \rfloor$ is a reasonable approximation for the number of clusters formed in the final opinion. However, when an agent's interaction range is limited to a topological neighborhood (neighbors defined by a network), more opinion clusters arise for smaller confidence-bound values.

In general, the agents' interactions occur along different social connections rather than a random selection across the entire population. Aside from the confidence bound, another condition of proximity is added to the DW model, i.e., agents only interact if they are directly related via a pre-existing social relation. The DW model is studied on different graph topologies such as a complete graph, square lattice, random graph (ER), and scale-free graph (BA) to understand the role of network topology in reaching consensus [29, 31]. The works demonstrate that for any $\epsilon > 0.5$, all the agents re-adjust their opinion and reach consensus to $x_i = 0.5 \, \forall i$ independent of the network topology, and smaller values of ϵ result in the formation of more clusters.

Deffuant et al. [32] extended the DW opinion dynamics model to examine the influence of

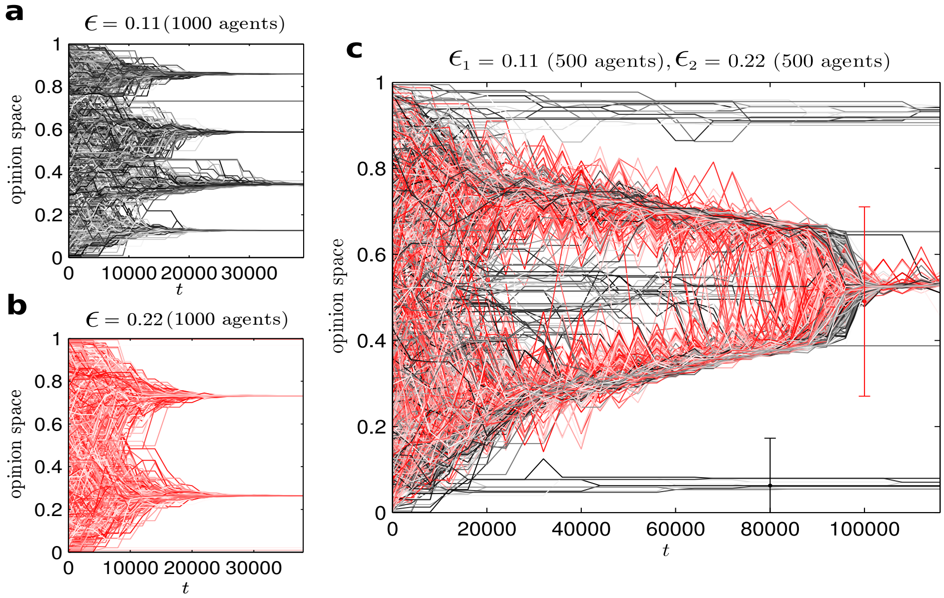


Figure 1.4: DW processes with 1000 agents. **a**: Closed-minded agents are black with $\epsilon = 0.11$ **b**: Open-minded in red with $\epsilon = 0.22$. **c**: Heterogeneous case with close-minded and open-minded agents. Initial conditions in all runs are equal. The agents can reach consensus under heterogeneous confidence bound but not in the corresponding homogeneous cases. *Figure and caption taken from Ref. [35]*

extremists without any specific interaction network. They showed that the extremists can have a very local influence or attract the whole population depending on the choice of parameters. Amblard et al. [33] later explored how social networks, particularly small-world networks with varying connectivity and randomness, affect this behavior. The shift to a single extreme opinion occurs beyond a specific connectivity threshold, which decreases with increasing randomness.

The DW model can also be defined in a discrete opinion space with $x_i(t) \in \{1, 2, \cdots, Q\}$ [34]. The agents interact with one another if their opinion difference is less than the confidence bound, i.e. $|x_i(t) - x_j(t)| < L$ where L is an integer instead of a real number. The system's dynamics remain the same following the Eq. (1.11) with the converge parameter μ. However, the opinion of agents after an interaction is rounded to the nearest integer. The original DW model can be recovered in the limit $L \to \infty$ and $Q \to \infty$ with a constant $\epsilon = L/Q$. If $L = 1, Q = 2$ on a complete graph, all the agents will reach the same opinion, which is the only stable point. Whereas for $Q > 2$, the dynamics will never reach a common consensus state and instead will have at least two clusters of opinion in the final opinion state.

The assumption that all agents apply the same criterion to determine whether to consider the perspectives of other agents is simplistic. When heterogeneity in confidence bound is introduced, several new traits emerge. The model is then commonly referred to as heterogeneous confidence bound model, i.e., agents can have different ϵ values [35]. To illustrate the problem, the dynamics are first evaluated with a homogeneous value of ϵ corresponding to close (small ϵ), open (large ϵ) minded agents, and then with heterogeneous value of ϵ, in which the population has an equal mix of close and open minded agents. From the numerical simulation, it is established that the system reaches a stable point with many clusters for a closed-minded society (see Fig. 1.4 **a**). For an open-minded society, the system stabilises to two opinions at the final state (see

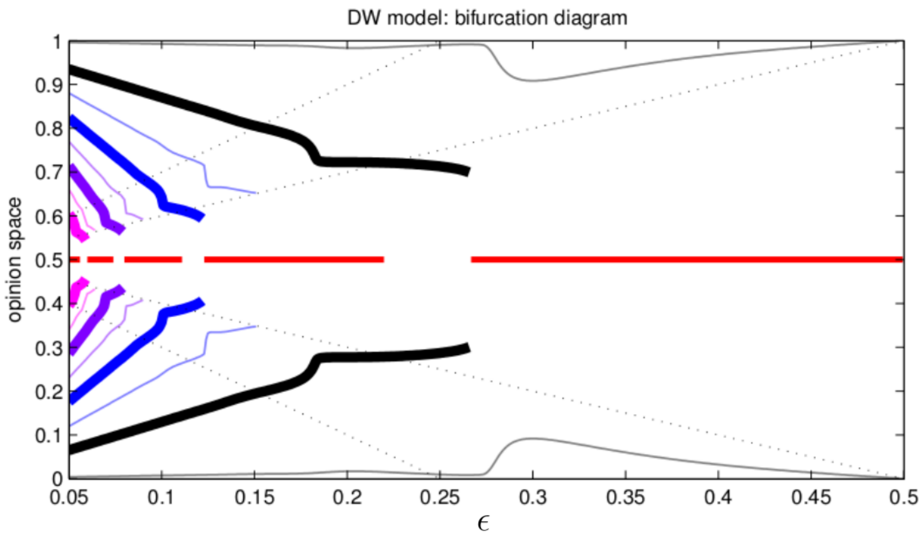


Figure 1.5: Bifurcation diagram of DW model. Exact location and masses of clusters in the DW model when cluster formation has stabilized, computed with interactive Markov chains with $S = [0, 1]$ divided into 1001 opinion classes and $\epsilon = [0.05, 0.5]$ and uniform initial distribution. The dotted lines represent the ϵ -interval around the central cluster. *Figure and caption taken from Ref. [37]*

Fig. 1.4 **b**). However, with a heterogeneous DW model, open-minded agents dominate the system dynamics (see Fig. 1.4 **c**).

Weisbuch [36] extended the model to use binary opinion vectors corresponding to m subjects, instead of scalars. Agents communicate based on a 'Hamming distance' metric, adjusting their opinions if they agree on at least $m - d$ subjects (where d is a confidence bound). The update rule is as follows: when opinions on a subject differ, a randomly chosen agent from the pair will be convinced by the other agent with a probability p. The results are consistent with the original Deffuant model, with p and N determining the convergence time. The key factors are d and the number of subjects m. Binary strings exhibit a abrupt phase transition from consensus to many clusters when d decreases.

The DW model can also be defined as density-based dynamics using a function that determines the density of agents in the opinion space [16, 38]. For the density-based dynamics model, the opinion of agents is considered as a density function in the opinion space $S \subset \mathbb{R}$ defined as $P(t, \cdot) : S \to \mathbb{R}_{\geq 0}$ with $\int_S P(t, x)dx = 1$ for all time t.

The evolution of the density function can be in continuous or discrete time. For a continuous time, the dynamics is defined using differential equation $\partial P(t, x)/\partial t = f(P(t, \cdot))$, where f operates on density function space. Then the differential equation is solved either analytically or numerically. The differential operator is replaced with the difference operator in the discrete-time case. Then we have $\Delta P(t + 1, x) = P(t + 1, x) - P(t, x) = f(P(t, \cdot))$ and the solution can be computed recursively from the equation $P(t + 1, x) = P(t, x) + \Delta P(t + 1)$. Numerical solutions are computed for both discrete [37] and continuous time cases [16, 38].

The Fig. 1.5 shows the bifurcation diagram of the homogeneous density-based DW model. Clusters are defined by collecting opinion space intervals where there is positive density. Therefore, one must choose a threshold at which mass is considered zero. A bifurcation diagram displays the locations of clusters in the limit density vs. the range of values of the confidence bound ϵ. So, for each confidence bound, one can identify the attractive cluster patterns and observe attractive pattern transitions at critical ϵ values. The locations of clusters with a major mass are indicated by fat lines, those with small masses by thin lines, and the central cluster by a medium line. The dotted lines are only for orientation; they represent the interval $[0.5 - \epsilon, 0.5 + \epsilon]$. A notable characteristic is the presence of minor clusters at the extremes and in between major clusters. Due to the small population sizes, these minor clusters do not usually evolve in agent-based simulations. For $\epsilon \geq 0.5$, the system reaches consensus, and there will be only one big cluster. As ϵ decreases, the system evolves into many clusters, as seen in agent-based DW models.

Hegselmann - Krause (HK) model

The HK model is similar to the DW model of opinion dynamics. However, in the HK model, a randomly selected agent, say i, updates its opinion by considering the opinions of all other agents within their confidence bound ϵ. Let there be N agents in the interaction network, and the opinion of each agent i is drawn following a uniform distribution within the interval $[0, 1]$. Then, we conduct systematic sweeps of the entire system to update each agent's opinion iteratively. The agent i can influence agent j, only if the difference between their opinion is less than the confidence bound ϵ. Suppose an agent i wants to update their opinion. In that case, one needs to determine which of their neighbors have their opinions within the confidence bound of agent i. All the neighbors of agent i within the interval $[x_i(t) - \epsilon, x_i(t) + \epsilon]$ can influence the opinion of agent i at time t, denoted as $N_i(t) = \{j : |x_i(t) - x_j(t)| < \epsilon, a_{ij} = 1\}$, where a_{ij} is the element of the adjacency matrix of the interaction network. The update rule is different in the HK model: instead of interacting with one of its compatible neighbors as in the DW model, agent i interacts with all its compatible neighbors simultaneously. During the dynamics, the opinion of agent i changes according to

$$x_i(t+1) = \frac{1}{|N_i(t)|} \sum_{j \in N_i(t)} x_j(t). \tag{1.12}$$

Repeating the process given by Eq. (1.12) will eventually lead to a configuration for the system which is a fixed point of the dynamics, making it stable. With several agents holding the same opinion, this configuration is characterized by few surviving opinions. It is proved in Ref. [39] that for a homogeneous HK model, the order of opinions remains unchanged by the dynamics, i.e.,

$$x_i(t) \leq x_j(t),\ \forall i \leq j,\ \Longrightarrow\ x_i(t+1) \leq x_j(t+1),\ \forall i \leq j. \tag{1.13}$$

Unlike DW model, where the convergence parameter must be specified, the confidence bound completely determines the HK model. In general, the main focus of the confidence bound models is to find the confidence bound ϵ at which the system reaches consensus and the time required to reach the consensus. The convergence time for a homogeneous HK model is proven to be in polynomial time regardless of the dimensions (i.e. for $x_i \in \mathbb{R}^d$). The authors of Ref. [40] established a lower bound of $O(N^2)$ and an upper bound of $O(N^3)$ for one-dimensional systems.

The HK dynamics evolve to the same stationary state pattern as in the DW model, with a decrease in the number of final opinion clusters as ϵ increases, as shown in Fig. 1.6. The *consensus threshold* ϵ_c is defined as the minimum value of confidence bound at which a network

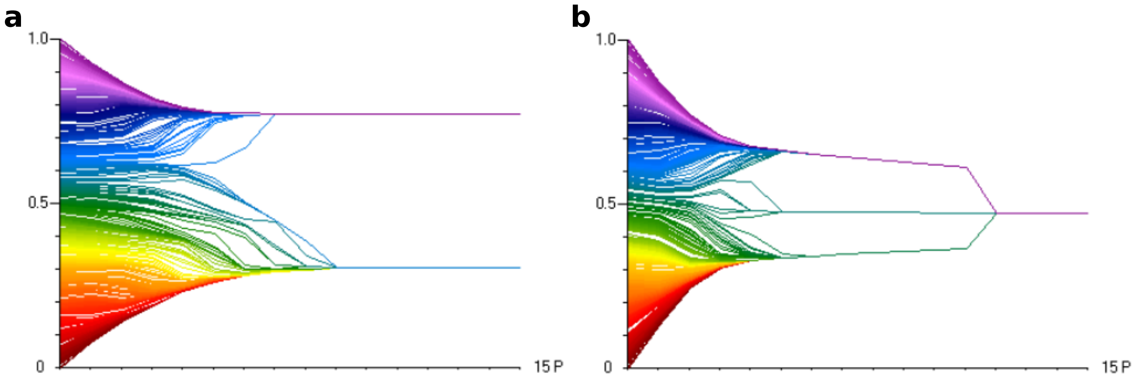


Figure 1.6: Time evolution of the HK model with 625 agents, randomly generated in the interval $[0, 1]$. **a**: $\epsilon = 0.15$, **b**: $\epsilon = 0.25$. *Figure and caption taken from Ref. [15].*

attains consensus. In particular, for $\epsilon > \epsilon_c$, all the agents re-adjust their opinion to form a single cluster (see Fig. 1.6 **b**). As the dynamics preserves the system's average opinion, as in the DW model, the final configurations on a complete graph are symmetric with respect to the central opinion $x_c = 0.5$ [6]. A detailed analysis of the clustering patterns as a function of confidence bound is performed in Ref. [41] and a subsequent study followed by modifying the agents' opinion range from $[0, 1]$ to $[0, L]$, with confidence bound $\epsilon = 1$ [42].

According to Fortunato [43], there can only be two values for ϵ_c, depending on the behaviour of the average degree d of the interaction network. This was studied on different types of graphs such as a complete graph, square lattice, scale-free graph, random graph, and star-like graph. When the population N goes to infinity, and if d diverges as $N \to \infty$, then the consensus threshold $\epsilon_c \sim 0.2$ on a complete graph. If d remains finite as $N \to \infty$, then the consensus threshold is $\epsilon_c = 0.5$, similar to the DW model.

The HK model is reformulated as an interactive Markov chains to shift the focus from a finite population of size N to a finite size opinion classes of size N_c [44]. Instead of using agents with a finite number of opinions, this enables the analysis of an infinite population spread over opinion classes. The dynamical behavior of the interactive Markov chain exhibits clustering and stability, similar to the agent-based models.

The assumption that all agents apply the same criterion to determine whether to consider the perspectives of other agents is very simplistic. When threshold heterogeneity is introduced, several new traits emerge [35]. To illustrate the problem, consider two values of ϵ corresponding to an open-minded and close-minded society. The dynamics are first evaluated with a homogeneous value of ϵ corresponding to close-minded (small values of ϵ), open-minded (large values of ϵ) agents and heterogeneous value of ϵ, in which the population has an equal mix of close and open minded agents. Through numerical simulation, it is established that the system reaches a stable point with many clusters for a closed-minded society (see Fig. 1.7**a**). For an open-minded society, the system stabilises to two opinions at the final state (see Fig. 1.7**b**). However, with a heterogeneous HK model, open-minded agents dominate the system dynamics (see Fig. 1.7 **c**). The open-minded agents influence the close-minded agents to reach a common consensus.

Pluchino et al. [45] expanded the HK model by introducing two-dimensional real-valued opinions, and by using Monte Carlo simulations on a complete graph for two different shapes of confidence bound ϵ (squared and circular). However, the number of clusters remains similar to the one-dimensional case. The consensus threshold was slightly higher for circular confidence

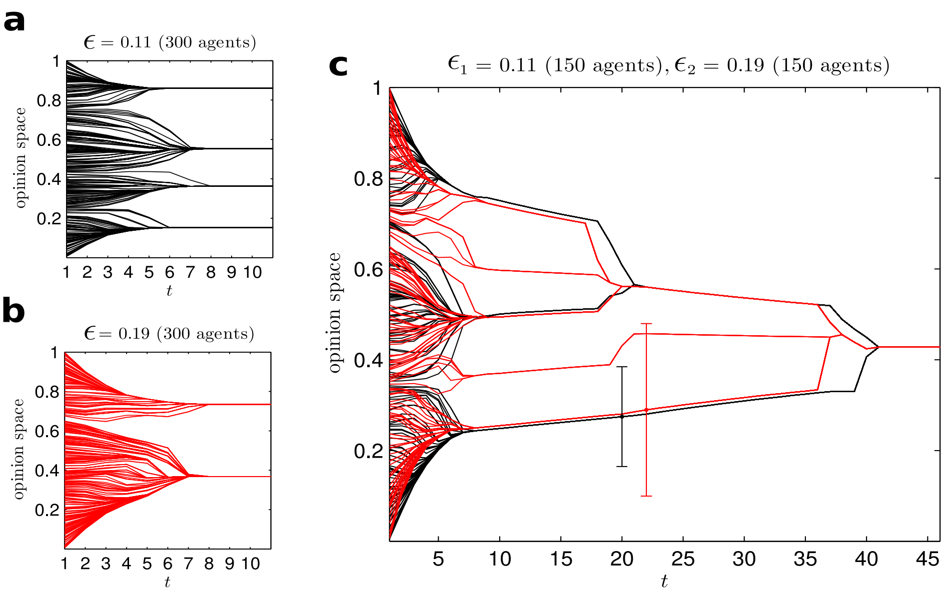


Figure 1.7: HK processes with 300 agents. **a**: Closed-minded agents are black with $\epsilon = 0.11$ **b**: Open-minded in red with $\epsilon = 0.19$. **c**: Heterogeneous case with close-minded and open-minded agents. Initial conditions in all runs are equal. The agents can reach consensus under heterogeneous confidence bound but not in the corresponding homogeneous cases. *Figure and caption taken from Ref. [35]*

bounds compared to the squared confidence bound due to the total area covered by the two. This study showed that in the continuum case with a large number of agents ($N \to \infty$), dynamics do not have much variation compared to the one-dimensional case. Other works on multi-dimensional HK model can be found in Refs. [46–48].

The HK model can also be defined as density-based dynamics for a density function that determines the density of agents in the opinion space [16, 37]. For the density-based HK dynamics model, the opinion of agents is considered as a density function in the opinion space $S \subset \mathbb{R}$ defined as $P(t, \cdot) : S \to \mathbb{R}_{\geq 0}$ with $\int_S P(t, x)dx = 1$ for all time t. For a given initial density function $P(0, \cdot)$, the dynamics of the density-based HK model evolves as a function in time. In order to define the density-based HK model, we need to define the $\epsilon-$ *local mean*,

$$M_1(x, P(t, \cdot), \epsilon) = \frac{\int_{x-\epsilon}^{x+\epsilon} yP(t, y)dy}{\int_{x-\epsilon}^{x+\epsilon} P(t, y)dy}. \tag{1.14}$$

This denotes the expected value of opinions in confidence bound interval $[x - \epsilon, x + \epsilon]$. Here $\int_{x-\epsilon}^{x+\epsilon} yP(t, y)dy$ denotes the first moment around x in the ϵ-interval and $\int_{x-\epsilon}^{x+\epsilon} P(t, y)dy$ provides necessary renormalization for the probability in that interval [16].

Fig. 1.8 shows the bifurcation diagram of the homogeneous density-based HK model. Clusters are defined by collecting opinion space intervals where there is positive density. Therefore, one must choose a threshold at which mass is considered zero. A bifurcation diagram displays the locations of clusters in the limit density vs. the range of values of the confidence bound ϵ. So, for each confidence bound, one can identify the attractive cluster patterns and observe attractive pattern transitions at critical ϵ values. The locations of clusters with a major mass are indicated by fat lines, those with small masses by thin lines, and the central cluster by a

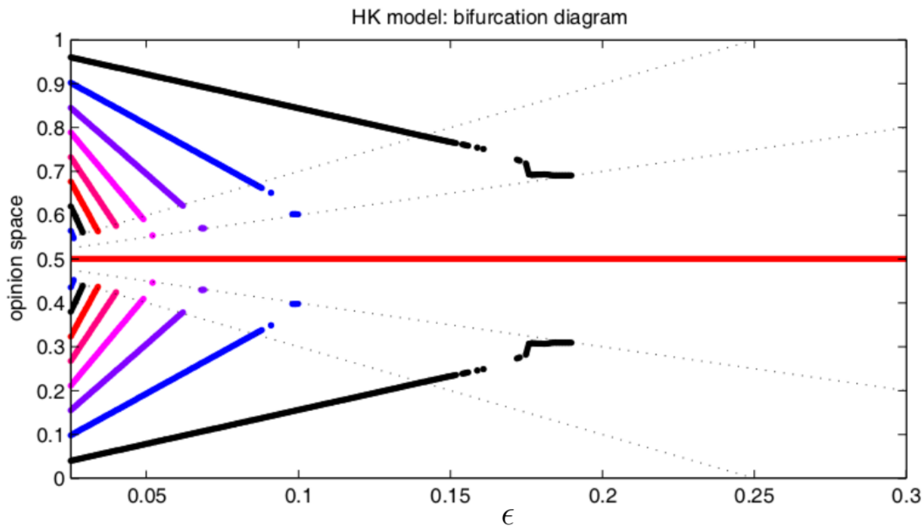


Figure 1.8: Bifurcation diagram of HK model. The exact location and masses of clusters in the DW model when cluster formation has stabilized, computed with interactive Markov chains with $S = [0, 1]$ divided into 1001 opinion classes and $\epsilon = [0.05, 0.3]$ and uniform initial distribution. The dotted lines stand for the ϵ -interval around the central cluster. *Figure and caption taken from Ref. [37]*

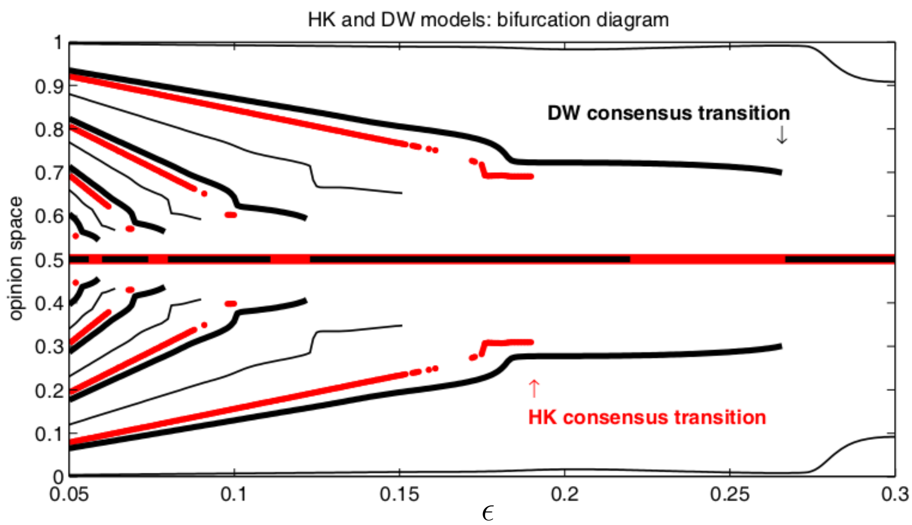


Figure 1.9: Bifurcation diagram for the density-based homogeneous DW and HK model. *Figure and caption taken from Ref. [37]*

medium line. In contrast to the DW model, the HK model does not contain any minor clusters. The system reaches consensus for $\epsilon \geq 0.19$, and is treated as the consensus threshold.

In conclusion, there are similarities and differences between DW and HK models. They mainly differ in terms of their convergence time and bifurcation diagram. Fig. 1.9 shows the bifurcation diagram of both DW and HK models for comparison. The critical value of ϵ for consensus transition differs for both models. Here, the consensus transition is the point of bifurcation between two major clusters and a central cluster, so the minor clusters in the DW model are neglected.

Generally, the confidence-bound models reach consensus, polarization or fragmentation with varying confidence bound ϵ values. However, these phenomena do not happen in the actual elections. In the real world, having two agents with the same opinion is impossible due to the complex nature of agents. The agents can agree on a specific issue, but this cannot be interpreted as two agents having the same opinion, as shown in the confidence-bound models or other models discussed above. Thus, more detailed constructions are required in the dynamics of agents' opinions and to convert them into the corresponding choices of candidates or parties when applying confidence-bound models to elections.

1.1.5 Voter Model

The voter model is considered one of the most appealing models of opinion dynamics [17]. It is also extensively used in probability theory and population genetics [6]. The voter model is defined on a set of N agents that hold one of two possible opinions, $x_i(t) = \pm 1$. At each time step t, an agent i is selected randomly along with one of its neighbors j. The agent i then updates its opinion as $x_i(t+1) = x_j(t)$ imitating its neighbor j. An illustration of voter model dynamics is shown in Fig. 1.10. The voter model was first introduced by Clifford and Sudbury [49] in 1973 and named 'voter model' by Holley and Liggett [50] because of its application to electoral processes.

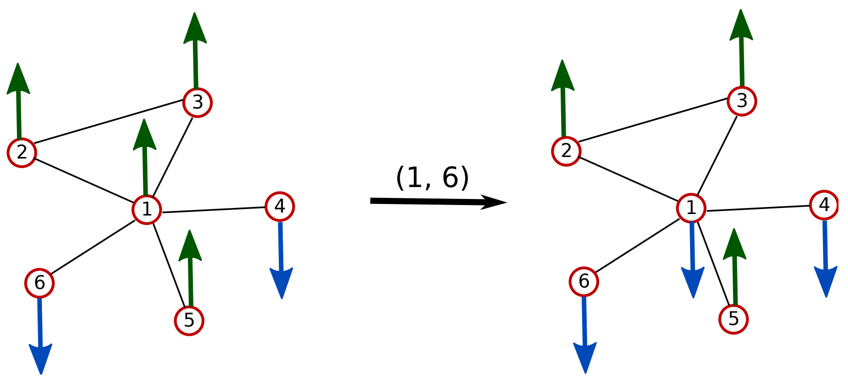


Figure 1.10: Illustration of the voter model. When agent '1' is selected along with one of its neighboring agent '6' with a different opinion as shown in the left, it adopts the opinion of agent '6', and the system evolves to the configuration shown on the right.

The voter model has been studied extensively when agents are modelled as vertices of $d-$ dimensional hyper-cubic lattices [8]. In a finite system, for any dimension d of the lattice, the voter dynamics reach consensus to one of the two possible outcomes: all the agents in the

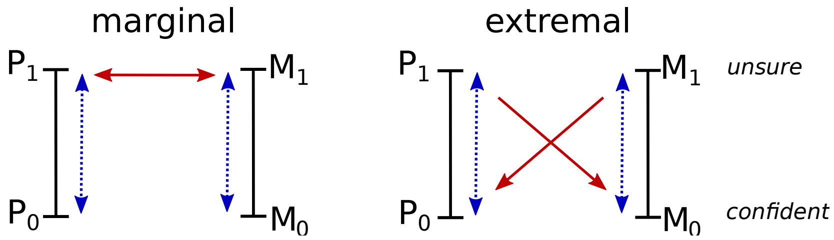


Figure 1.11: Illustration of the possible states of agents (P_0, P_1, M_0, M_1) and their transitions. Dashed arrows indicate possible confidence level changes (biased towards higher confidence), while solid arrows indicate possible opinion change events. *Figure and caption taken from Ref. [52]*

population with the same opinion of either $+1$ or -1. The probability of reaching consensus to one or the other possible opinions depends on the initial state of the agents. It is proved in Ref. [51] that the time T_N needed to reach consensus depends on the system size N: $T_N \sim N^2$ for $d = 1$, $T_N \sim N \ln N$ for $d = 2$ and $T_N \sim N$ for $d > 2$. For an infinite system, a consensus state is reached only for smaller dimensions, $d \leq 2$.

The lack of self-confidence of an agent is one of the many unrealistic assumptions in the voter model. There are different ways to model how much weight is assigned to the opinion of other agents compared to their own self-confidence to decide what to believe and what not to believe. This could be incorporated in the voter model by considering that the agents would require multiple encounters with neighbors with different opinions before the agent actually changes his opinion.

One such model is the *confident voter model* with two additional levels of commitment - confident and unsure [52]. In this model, positive (P) and negative (M) denote the two opinion states and subscript '0' denotes the confident agents whereas the subscript '1' denotes the unsure agents. Thus the possible states of an agent are P_0, P_1, M_0 and M_1. When a confident voter interacts with an agent of a different opinion, the confident voter becomes less confident or unsure but does not change their opinion, while an unsure agent changes their opinion.

There are two variants of confident voter model named 'marginal' and 'extremal'. In the case of 'marginal' voting, unsure agents who change their opinion are also unsure about their new opinion. However, in the case of 'extremal' voting, the unsure agents gain some confidence after changing their opinion and need two interactions with agents of the opposite opinion to change their opinion again. The assumptions and dynamics for marginal and extremal cases of confident voter models are sketched in Fig. 1.11.

The assumption described in discrete time can be conveniently translated to a mean field description using the corresponding densities of the four types of agents. Then the agent type and their densities can be denoted as P_0, P_1, M_0 and M_1. The dynamics of 'marginal' version for positive agents can be written as

$$\begin{aligned} \dot{P}_0 &= -(M_0 + M_1)P_0 + P_0P_1 = MP_0 + P_0P_1, \\ \dot{P}_1 &= MP_0 - P_0P_1 + M_1P_0 - M_0P_1. \end{aligned} \tag{1.15}$$

The corresponding equation for M is obtained by interchanging $M \leftrightarrow P$. The rate equation for the total density of positive agents is

$$\dot{P} = M_1P_0 - M_0P_1, \tag{1.16}$$

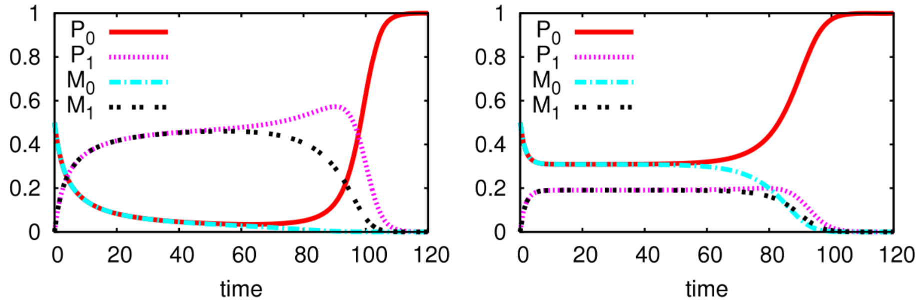


Figure 1.12: Evolution of the densities for the: (left) marginal and (right) extremal models with the near-symmetric initial condition $P_0 = 0.50001$, $M_0 = 0.49999$ and $P_1 = M_1 = 0$. *Figure and caption taken from Ref. [52]*

and from the complementary equations for $\dot{M}$, we verify that the total density of agents is conserved, i.e. $\dot{P} + \dot{M} = 0$. Now, for the 'extremal' version, the rate equations are

$$\begin{aligned} \dot{P}_0 =& - M_0P_0 + M_1P_1 + P_0P_1, \\ \dot{P}_1 =& M_0P_0 - M_1P_1 - P_0P_1 + (M_1P_0 - M_0P_1), \end{aligned} \tag{1.17}$$

and the corresponding equations for M can be obtained by interchanging $M \leftrightarrow P$. The rate equation for the total density of positive agents is the same as that for the marginal version, and the total density of agents is conserved, $\dot{P} + \dot{M} = 0$. The time evolution of both model variants is shown in Fig. 1.12. On a complete graph, the confident voter model converges to confident consensus in a time that grows as $\ln N$ after crossing a mixed state of unsure agents. However, on a regular lattice, the consensus time grows as a power law in N, with some configurations crossing a long-lived striped state.

Another possible extension of the model is the introduction of zealots [53]. When the agents are placed on hypercubic lattice of one and two-dimensions, a single zealot influences infinite group of agents in the system to reach a consensus with its opinion due to the recurrent interactions between the agents. However, at higher dimension, the random interactions are transient, and therefore there is a finite probability that agents never interact with the zealots. Thus the zealot could influence only the agents in its neighborhood and cannot influence the agents that are farther away. Thus, the system will not reach a consensus in higher dimensions. In a finite system, when a small equal number of zealots are added for each opinion ($Z = 2$ for 1000 agents, where Z is the number of zealots), zealots have been shown to prevent consensus or form robust majorities [54]. The system shows a Gaussian distribution of the average opinion of agents in one and two dimensions with a width proportional to $1/\sqrt{Z}$, and independent of the number of agents in the system.

The ultimate fate of an agent's opinion is analyzed in a population comprising leftists, centrists, and rightists who socially interact with one another [55]. A centrist and a leftist can become both centrists or leftists with equal rates after the interaction between agents (and similarly for a centrist and a rightist). Leftists and rightists, however, do not interact. The interaction process is continued until the system reaches an equilibrium or stable state. The authors determined the probability of system attaining consensus (either leftist, rightist, or centrist) or a frozen mixture of leftists and rightists in the mean-field limit as a function of the initial composition of the population along with the average time needed to achieve the final state.

Mobilia [56] proposed similar model with three opinions; leftist, centrist and rightist on a complete graph. A parameter q quantifies the bias towards extremism ($q > 0$) or centrism ($q < 0$), and governs transitions from an extreme to a centrist opinion. In a population of size N, a leftist and centrist can both become leftists with rate $(1 + q)/2$ or centrist with rate $(1 - q)/2$, and similarly for rightist and centrist. The authors demonstrate that polarization is preferred for $q > 0$; nonetheless, a finite probability for consensus always exists. Whereas, for $q < 0$, the system has higher probability to reach consensus. The system's behavior is altered by adding centrist zealots (centrists who maintain their position) [57]. A large fraction of centrists zealots $Z_c > \frac{q}{1+q}$ generates consensus on the centrist opinion in the system. Populations with a small proportion of zealots tend to be mixed, with centrists coexisting with either both extremist type (where the two are equally convincing) or with the most convincing one.

Castellano et al. [58] introduced the non-linear q-voter model, a generalization of discrete opinion models. In this model, N individuals are placed on the vertices of a complete graph and hold an opinion ± 1. At each time step, a set of q agents are chosen at random, and they form a group. If all the agents in the group have the same opinion, they influence one neighbor chosen at random, i.e. this agent adopt the opinion of the group. If the group disagrees, the chosen neighbor flips its opinion with probability p. Analytic results for $q \leq 3$ confirm the numerical results, indicating a transition from an ordered phase for small values of p to a disordered one for large values of p. However, for $q > 3$ in the q-voter model, a novel form of transition between the two phases is observed. This transition involves an intermediate regime where the final state of the model is dependent on its initial conditions. Researchers have also investigated the behavior of q-voter model on random regular networks. According to Moretti et al. [59], the intermediate regime disappears, resulting in a behavior similar to that observed on a lattice.

There are always some agents that go against the majority opinion, named anti-conformity agents. Nyczka et al. [60] conducted an analysis of the q-voter model to compare the dynamics of non-conformity and anti-conformity. Non-conformity refers to the behavior of some agents who flip their opinion with probability p, irrespective of the group's opinion. On the other hand, anti-conformity refers to some agents who oppose the group's opinion and choose the opposite one with probability p. Although these two dynamics seem similar, the analysis revealed significant differences between them. For instance, the critical value p_c for the order-disorder phase transition is shown to increase with q in the case of anti-conformity, whereas it decreases with q for non-conformity.

1.1.6 The Majority Rule Model

The Majority Rule (MR) model was proposed to model public debates in which each agent has information about the opinions of all other agents [61]. There are $N_{+1}(t)$ agents with positive and $N_{-1}(t) = N - N_{+1}(t)$ agents with negative opinion. The probability that an agent has positive opinion is given by $P_{+1}(t) = N_{+1}(t)/N$, while the probability to have negative opinion is given by $P_{-1}(t) = 1 - P_{+1}(t)$. From a random initial opinion configuration of the agents, agents starts discussing in small groups. However they don't meet all the time and all together at once. The probability that a group of size g is selected for discussion is given by a_g, which satisfy the constraint $\sum_g a_g = 1$ where $g = 1, 2, \cdots, G$. Within each group, all agents adopt the majority opinion within the group, as shown in Fig. 1.13.

During the discussion, all agents are assumed to be involved in one group, i.e. the groups do not overlap at any time t. This means that a given agent, on average, is a member of a group of size g with probability a_g. The discussions in groups occur sequentially in time, and each

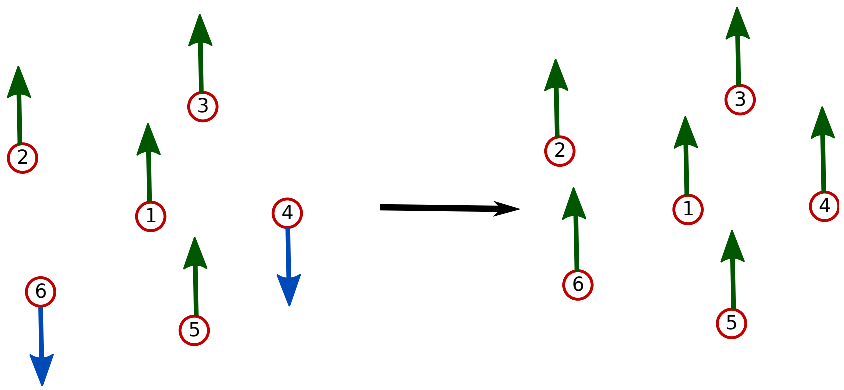


Figure 1.13: Illustration of Majority model for one random time step t with a group of $g = 6$ agents selected randomly. The image on the left denotes the opinion of agents before the discussion and, on the right, the opinion of agents after the discussion.

time with different groups. Each new cycle of multi-size discussions is assumed to be a time increment of 1.

The randomly selected group for discussion can be of odd or even size. If the group is odd, there is an absolute majority, and the majority opinion is adopted among the agents. In contrast, if the group is even, there may be a tie between the opinions. In that case, one way of selecting a majority opinion is by imposing a bias in favor of one opinion. In Ref [61], when a tie appears, the dynamics favours the negative opinion. The probability that an agent will have a positive opinion at any time $t+1$ depends on probability that the agent belongs to the group of size g given by a_g and the majority opinion within the group. Then the recursion relation on the probability of agents can be written as [61]

$$\begin{aligned} P_{+1}(t+1) &= \sum_{g=1}^{G} a_g \sum_{j=[\frac{g}{2}+1]}^{g} C_j^g P_{+1}(t)^j P_{-1}(t)^{g-j}, \\ P_{-1}(t+1) &= \sum_{g=1}^{G} a_g \sum_{j=[\frac{g}{2}]}^{g} C_j^g P_{-1}(t)^j P_{+1}(t)^{g-j}, \end{aligned} \tag{1.18}$$

where $C_j^k = \frac{k!}{(k-j)!j!}$ and $[x]$ denotes the integer part of x. This equation conserves the total probability $P_{+1}(t+1) + P_{-1}(t+1) = 1$ at any time t. This can be shown by changing the dummy variable j to l by substituting $l = g - j$ for $P_{-1}(t+1)$ in Eq.(1.18), and adding it with $P_{+1}(t+1)$. It is to be noted that Eq. (1.18) is a mean-field equation that neglects possible fluctuations.

The dynamics of MR model have two stable fixed point $P_{+1} = 0$ and $P_{+1} = 1$, and an unstable one, called the faith point at $P_{+1} = p_c$. The unstable fixed point p_c is located in between the stable fixed points and is given by [62],

$$\lim_{t\to\infty} P_{+1}(t) = \begin{cases} 0 & \text{if} \quad P_{+1}(0) < p_c, \\ 1 & \text{if} \quad P_{+1}(0) > p_c. \end{cases} \tag{1.19}$$

As Eq. (1.18) neglects possible system size fluctuations, the result in Eq.(1.19) is only valid in the limit $N \to \infty$. The time to reach a consensus scales to the order of log N number of updates per agent. In addition, the value of p_c depends on $\{a_g\}$, G and the initial value $P_{+1}(0)$. If the size of the group is even, $p_c < 1/2$, i.e. consensus will be reached for negative opinion even if it is initially held by a minority of the agents. Whereas if the group size is odd, $p_c = 1/2$, due to the symmetry of the two opinions [8, 63].

Galam et. al [64] introduced another method of selecting a majority in case of a tie by introducing a probability p by which the agents in the group adopt the positive opinions, and with $(1 - p)$ the negative opinions. With such an extension, authors studied the emergence of democratic or dictatorial extremism. In this study, the agents are assumed to belong to a specific social-cultural class. They are randomly assigned to various groups within their respective class at each step to evolve locally following the majority rule. The value of p represents the average individual biases caused by the presence of heterogeneous views in the corresponding class, which differs from class to class. The approach leads to extremism, with each class completely polarized along one opinion. For homogeneous classes ($p = 0$), extremism may follow the initial minority, making it minority induced. Contrarily, heterogeneous classes ($p = 1/2$) show more balanced dynamics, which leads to extremism brought on by the majority. Segregation within sub-classes may result in the coexistence of opinions at the class level, thus preventing global extremism.

A natural extension of the model is to incorporate a small fraction of zealots [63, 65]. The existence of zealots for only one of the two opinions shifts p_c to a value lower than 1/2 in favor of the zealots's side. Furthermore, a fraction of zealots above 0.17 ensures that their side wins regardless of the initial conditions. Additionally, balanced dynamics are restored by an equal number of zealots on both sides (i.e., $p_c = 1/2$). A similar extension of the MR model is used to explain the result of public debates driven by incomplete scientific data [66]. When scientific data is insufficient, the zealots drive the results of the debate. Therefore, to win a debate, one strategy is to increase the proportion of zealots.

Understanding the consequences of free will and the collective opinions of agents is important to understand how opinion dynamics can be leveraged to shape public opinion. To address this problem, an extension of the MR model was developed by Wu et al. [67], which introduces two distinct models (model I and model II) that feature various dynamics of opinion-updating. While updating opinions, the agents interact with their neighbors according to the majority rule with probability $1 - p$, and with probability p they can exercise their own free will and make decisions (model I) or follow the mass opinion (model II). They showed that for both models, there exist a threshold for p below which consensus is reached.

The extensions of MR model are employed to investigate how the interplay between individual obstinacy and collective beliefs influences the course of public debates [68], the formation of bullish trends driving an asset price in market [69]. A stochastic version of this model on hypergraphs is used to understand how opinions are formed and how they evolve over time in more realistic and complex network structures [70]. The non-consensus opinion (NCO) models (i.e. models in which two opinions coexist in the stable state above a certain threshold of interactions), and its extensions are one class of models that resemble the MR model [71]. These models introduce self-opinion into the majority rule, either with or without weight. It is demonstrated that the system can reach stable states where the two opposing opinions coexist on many types of networks, including fully-connected networks. Galam [72] contains a comprehensive examination of the MR model's extensions and applications.

1.1.7 Sznajd Model

The Sznajd model (SM) was introduced in 2000 as a sociophysics model by Katarzyna Sznajd-Weron based on the maxim of trade union 'United we Stand, Divided we Fall' [73]. The model later came to be known as the 'Sznajd model' due to Dietrich Stauffer [74]. Sznajd's study concentrated on how opinions are spread in a society with the agents assumed to be placed on a one-dimensional chain [73]. The opinion of agents is assumed to take one of the two opinions $x_i = \pm 1$ (usually referred as up or down spin). Motivated by the phenomenon of *social validation*, the dynamics of the Sznajd model is defned as follows,

1. At each time step, a pair of agents x_i and x_{i+1} are chosen at random to change the opinion of their neighbors x_{i-1} and x_{i+2}.

2. If $x_i = x_{i+1}$, then $x_{i-1} = x_i$ and $x_{i+2} = x_{i+1}$.

3. If $x_i = -x_{i+1}$, then $x_{i-1} = x_{i+1}$ and $x_{i+2} = x_i$.

Opinions of the agents are updated in a random sequential order. Since the introduction of Sznajd model, it has been argued that one of its distinguishing features is the information flow from the initial pair of agents to their neighbors, and thus the dynamics is called 'outflow-dynamics' [6, 75]. In other opinion dynamics models, agents are instead influenced by their neighbors.

The *decision time* of an individual is defined as the time needed for an individual to change their opinion. Using standard Monte Carlo simulations, it is demonstrated that simple rules of Sznajd model results in quite sophisticated dynamics, and a power law with the exponent $\tau \sim 1.5$ for the decision time [73]. Starting from a random initial configurations where both the opinions are equally distributed, it has been shown that the system evolves into either a full consensus to $+1$ or -1 with a probability 0.25 each or a stalemate state (inability to have a common opinion) with a probability 0.5 [73]. The SM model has been extended and applied in marketing [76], finance [77], and politics [78].

The Sznajd model has been modified by Schulze [76], by changing the dynamic rules for the condition $x_i = -x_{i+1}$ to,

$3'$. If $x_i = -x_{i+1}$, then $x_{i-1} = x_i$ and $x_{i+2} = x_{i+1}$.

With the new dynamic rules, the system always reaches full consensus to either ± 1 with a probability of 0.5 each by avoiding the stalemate state. Other variations of the model to avoid the stalemate final state are discussed in Ref. [73]. These new rules were called 'if you do not know what to do, do nothing' and 'if you do not know what to do, do whatever'.

Behera et al. [79] compared the the 'voter model' and 'Sznajd model'. They showed that the dynamics of a one-dimensional voter model and the original Sznajd model are equivalent. The only difference is, in voter model the agent is influenced by it's nearest neighbor whereas in Sznajd model the agent is influenced by the next-to-nearest neighbor. The fact that the Sznajd model updates the opinion of two agents at a time, whereas in voter model only one agent updates it opinion introduces a factor of two in the average time to reach stable state. All other features remain the same, including the decision time.

Stauffer [80] extended the original Sznajd model to two-dimensional square lattice with simultaneous updating rules (agents receiving simultaneous information from different sources). A pair of neighboring agents on a square lattice can persuade their six neighbors to agree with

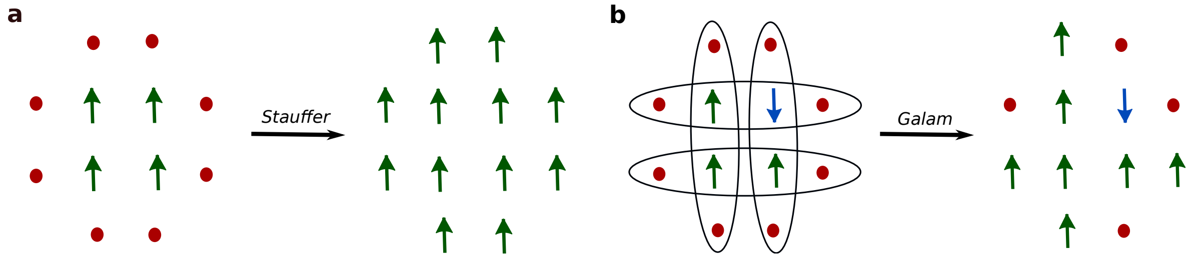


Figure 1.14: Dynamical rules of Sznajd model on the square lattice proposed by Stauffer and Galam in private communication described in Ref. [81]. In every time step a set of four opinions denoted by arrows is picked at random and influences opinions denoted by dots producing the configurations presented in the right-hand side pictures. Dots represent opinions, which can be either ± 1. *Figure and caption taken from Ref. [74]*

them if and only if they both hold the same opinion. It is found that it is far more difficult to reach a complete consensus with simultaneous updating rules. This is because some agents refuse to adjust their opinions when they simultaneously receive conflicting information from several neighbor pairs (a phenomenon known as frustration).

Stauffer et al. [81] suggested several ways to generalize Sznajd model to square lattices. They displayed the model on an $L \times L$ square lattice, $x_i = \pm 1$. Six new rules were introduced, but only two of the six introduced rules have been applied in subsequent publications:

1. Four neighbors on a square lattice leaves its eight neighbors unchanged if any one of the four neighbors have a different opinion (see Fig.1.14 **a**).

2. A neighboring pair persuades its six neighboring agents to follow their orientation if and only if the two pair opinions are same.

With both these rules, the system always reach complete consensus as a steady state. Furthermore, a phase transition is observed, where initial densities of positive opinion $P_{+1} < 1/2$ leads to all negative opinions and $P_{+1} > 1/2$ to all positive opinions for large enough systems.

Galam in a private communication with Stauffer (as mentioned in Ref. [81]) showed that the updating rule of the one-dimensional Sznajd model can be precisely translated into two dimensions as shown in Fig. 1.14 **b**. The one-dimensional rule is applied to each of the four chains of four opinions that are centered about two horizontal and two vertical pairs.

The Sznajd model is as well extended to different complex networks, including scale-free networks [34], small-world network [82] and complete graph [83]. The numerical results on complete graph are supplemented with analytical results and proved the existence of a phase transition in the original formulation of Sznajd model [83]. More complex topologies, such as two-dimensional lattices or networks, has undeniably brought Sznajd model closer to reality.

Kondrat et al. [84] conducted a recent study on disagreement in the Sznajd model in one dimension, where both conformist (agreement) and anti-conformist (disagreement) behaviors are observed. A new parameter called p is added to the model specifically to describe the likelihood that, in a situation where two neighbors have the same opinion, a third neighbor—who previously shared that opinion—will adopt the opposing opinion. As in the original Sznajd model, if the third neighbor did not share the initial pair's opinion, then the neighbor adapt the opinion of the pair. It is demonstrated that when anti-conformity is low, consensus can be achieved and population-wide spontaneous changes between ± 1 occur. On the other hand side, excessive anti-conformity causes the average opinion level to oscillate around 0 without reaching 1.

The model has also been employed in a study that compares two opinions, one based on the Sznajd model and the other on the voter model [85]. Agents are assumed to be placed on the vertices of Watts-Strogatz small-world network and can take one of two states, S or D. An agent is chosen at random for each time step. If the selected agent is in state S, it randomly switches one of its neighbors from state D to state S. Whereas if it is in state D, then one of its neighbors (if exists) is selected at random with a probability p and transforms all of its neighbors into state D. When the clustering coefficient is low, it is demonstrated that the system switches from complete consensus on S to complete consensus on D depending on p. However, the opinion S becomes more prevalent as the clustering coefficient increases.

Sznajd [75] extended the model to investigate the effect of two types of social influence: conformity and independence. The outflow dynamics describe the conformity in the model as in the original Sznajd model. Independence means that a neighboring agent has a probability p of choosing not to follow an agreement plaquette. In this case, they have the agent flexibility to change their minds with probability f. The model is examined on a complete graph, one-dimensional, and two-dimensional lattices. The majority is larger for low levels of independence (p), which encourages the coexistence of the two opinions in society.

1.1.8 Multidimensional Models

In most of the models discussed before, the opinion of the agents are described by scalars. Agents can have opinion on various topics, including sports, entertainment, spirituality, and so on, which can have significant influence during the interactions among their neighbors. In order to accommodate opinions on different topics simultaneously, several multidimensional opinion dynamics models have been put forth.

One of the first opinion dynamics models that considers a multi-dimensional opinion is Axelrod [86] model. The agents are assumed to be located on the nodes of a network or on the sites of regular lattice. The model represents the culture of an individual within a population of size N using F variables $(\sigma_1, \cdots, \sigma_F)$, where each variable can take one among q discrete values, $\sigma_f = 0, 1, \cdots, q-1$. These variables, called cultural *features*, represent an individual's beliefs, attitudes, or behavior, and q is the number of possible *traits* allowed per feature. At time t, the interactions between agent i and one of its neighbor j occur based on the extent to which their cultural features *overlap*:

$$o_{ij} = \frac{1}{F} \sum_{f=1}^{F} \delta_{\sigma_f(i), \sigma_f(j)}, \tag{1.20}$$

where δ_{ij} represents Kronecker's delta.

With a probability o_{ij}, there is an interaction between agents i and j. If they interact, one feature with different traits is selected $(\sigma_f(i) \neq \sigma_f(j))$ and is set to $\sigma_f(i) = \sigma_f(j)$; otherwise nothing happens. The dynamics follows a random sequential order. The Axelrod dynamics tends to make the interacting agents more similar, and the interactions are more likely among the agents that already shares many traits. Whereas agents with no shared traits cannot interact with each other.

The system evolves to one of many possible final states from a finite random initial population. These final states fall into two types: ordered states, where all agents share the same traits

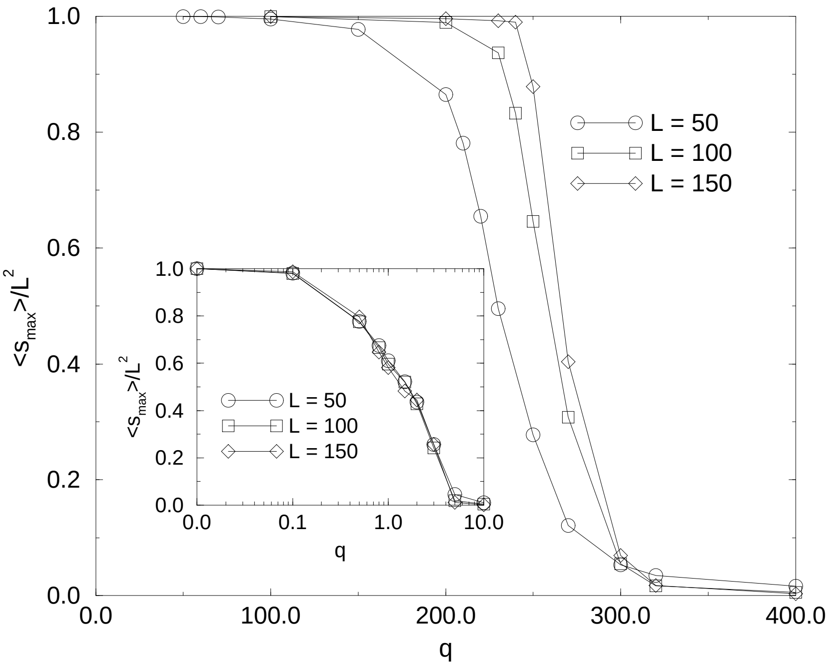


Figure 1.15: Behaviour of the order parameter $\langle S_{max} \rangle / L^2$ vs. q for three different system sizes and $F = 10$ in Axelrod model on a regular lattice of size L. In the inset the same quantity is reported for $F = 2$. *Figure and caption taken from Ref. [87]*

.

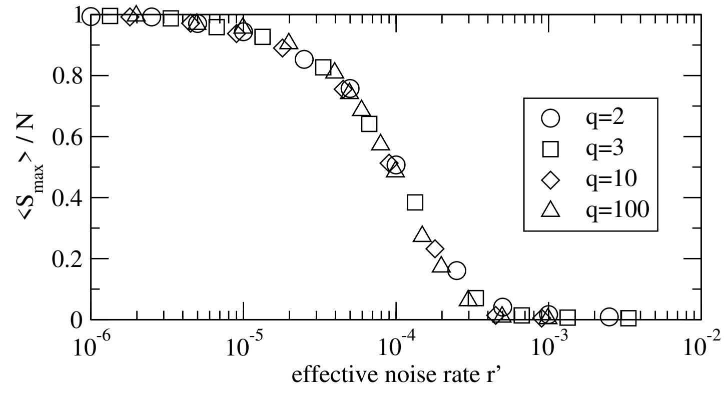


Figure 1.16: Order parameter $\langle S_{max} \rangle / N$ as a function of the effective noise $r' = r(1 - 1/q)$ for different values of q in Axelrod model. Simulations have been run in systems of size $N = 50^2$ with $F = 10$. *Figure and caption taken from Ref. [88]*

(q^F possible states), and frozen states, where multiple cultural regions coexist. The final state can be defined by the distribution of cultural region sizes, where a region is defined as the connected set of sites sharing exactly the same features, and $\langle S_{max} \rangle$ denotes the average size over the largest regions obtained from different realization of random population. The number of possible traits q in the initial population determines which type of final state is reached [87]. If q is small, there are many shared traits among agents, which facilitates interaction and increases similarity, leading to consensus. For larger q, there are only a limited number of shared initial traits, resulting in limited interaction and the formation of cultural domains that are unable to grow, preventing the consensus from being reached. When the agents are placed at the vertices of regular lattices of size L, the phase transition between the two types of states occurs at a critical value q_c, which depends on F (see Fig. 1.15).

There are various ways to define the order parameters to describe the transition, including the average fraction $\langle S_{max} \rangle / N$ of the system occupied by the largest cultural region. In the ordered phase, this fraction is finite (as N approaches infinity). On the other hand, in the disordered phase, cultural domains are of finite size, and therefore $\langle S_{max} \rangle / N$ scales as $1/N$. Another commonly utilized order parameter has been introduced by González-Avella et al. [89], $g = \langle N_g \rangle / N$, with N_g denoting the number of different regions present in the final state. During the ordered phase, g approaches zero, whereas it remains finite in the disordered phase.

The Axelrod model has been extensively studied since its inception. Although numerical studies dominate the research, Lanchier [90] presented some analytical proofs, demonstrating that when $F = q = 2$, the majority of the population forms one cluster, whereas a partial proof shows that if $q > F$, the population remains fragmented. Lanchier and Scarlatos [91] explored the one-dimensional case and found that the system fixates (each individual's state gets updated a finite number of times before the system reaches stable state) when $F \leq cq$ where $e^{-c} = c$. In Kempe et al. [92], a similar model, which generalizes the use of homophily and influence in models, was introduced. They proved analytically that when all agents interact, all initial conditions lead to convergence to a stable state.

Barbosa and Fontanari [93] analyzed the relationship between the number of cultural clusters and lattice area ($A = L^2$, where L is the size of the lattice). They discovered that when $F \geq 3$ and $q < q_c$, there is a strange non-monotonic relationship between the number of clusters and A. In particular, above a specific threshold of the area A, the number of coexisting clusters decreases, in contrast to the well-known results for species-area relation (monotonical increase of the number of species with the size of the area of a particular habitat). However, when outside these parameter values, the expected culture-area relation (relation between the number of coexisting cultures and the area) is observed, described by a curve that initially increases linearly with A, then flattens when the maximum number of possible clusters (q^F) is reached.

The ability of an individual to change one of their traits spontaneously, regardless of their surroundings, is referred to as 'cultural drift' in social science. This phenomenon corresponds to the introduction of flipping events driven by random noise. Klemm et al. [88] showed that incorporating noise at a rate r, results in a noise-induced order-disorder transition, which is independent of the value of parameter q (see Fig. 1.16).

The Deffuant model paper [29] introduced a modification to vectorial opinions, where binary variables were used instead of continuous ones. This resulted in a model resembling the Axelrod model but with a maximum of two traits per feature ($q = 2$) and a probability of interaction determined by a step function at a fixed threshold ϵ, representing confidence bound. Mean field analysis revealed a transition from full consensus to fragmentation as ϵ decreases.

Laguna et al. [94] explored a similar model in which the probability of two agents interacting and each pair of different variables becoming equal is determined by a probability p when the distance between the opinion of agents is less than the confidence bound ϵ. The properties of the transition between consensus and fragmentation vary depending on whether $p = 1$ or $p < 1$, but are still determined by the confidence bound ϵ.

Jacobmeier [95] conducted a study to simulate a Deffaunt consensus model on a directed Barabási-Albert network using a multi-component subject vector. The study aimed to explore the distribution and clusters of agents who agree on opinions regarding subjects. The findings suggest that there is no absolute consensus, and the existence of a robust cluster depends on the number of subjects. The study also found that two agents tend to agree either on all or no subjects, and the tolerance in the change of opinions of group members is a crucial parameter that affects the consensus-forming process.

Huet [96] investigated the dynamics of opinion change in two dimensions that occur through social interaction. The authors introduce a rejection mechanism into the two-dimensional bounded confidence model proposed by Deffuant et al. [29]. The agents are characterized by continuous opinions associated with constant confidence bound ϵ. The agents interact randomly in pairs, and if their opinions are within ϵ on both dimensions, the bounded confidence model applies. However, if their opinions are within ϵ on one dimension and beyond $\epsilon + \delta\epsilon$ on the other dimension, they are in a dissonant state, and they move away from their closer opinions to solve the issue. The model reveals the presence of metastable clusters maintained by competitor clusters' opposing influences. The study indicates that the number of clusters increases linearly with the inverse of confidence bound for a large range of confidence bound values, whereas it grows quadratically in the bounded confidence model. The analysis and initial experiments support this hypothesis.

Lorenz [97] extended Deffuant-Weisbuch and Hegselmann-Krause to multidimensional continuous opinion dynamics, where opinions are nonnegative vectors of size k whose components sum up to one. The model employs bounded confidence using the Euclidean distance between two agents, denoted as d_{ij}. The model converges to one or more clusters, depending on the values of ϵ and k. When k increases, the model leads to better chances for a vast majority consensus even for lower confidence bound ϵ, but the number of minority clusters rises with k, too. Additionally, if ϵ exceeds a threshold, the population converges to a single opinion, which decreases with the number of options k. Ref. [98] offers a comparison between this approach and the approach of considering the k options as independent. In the case of independent options, the increase of k does not lead to an improvement in agreement.

Etesami et al. [99] examined the behavior of the Hegselmann-Krause model of opinion dynamics under different conditions. They investigated the termination time of the Hegselmann-Krause dynamics in arbitrary finite dimensions and concluded that the termination time depends only on the number of agents involved in the dynamics. They argue that it is the most precise limit found to date, which eliminates the dependency of the termination time on the ambient space's dimension.

Most of the studies on multidimensional opinion dynamics models concentrates on the scenario of two or three different opinion types. Waagen [100] generalizes this into k general opinions along with zealots. The study discusses opinion dynamics models in the mean-field case of the naming game model, where an arbitrary number of opinions are considered. The naming game model assumes an infinite population on a complete graph. The article presents a method for generating mean field dynamical equations for the general case of k opinions. The steady states are calculated in two special cases: the case where there exist zealots of only one type and the case where there are an equal number of zealots for each opinion. The article shows that in these special cases, a phase transition occurs at critical values p_c of the parameter p describing the fraction of zealots. In the former case, the critical value determines the threshold value beyond which it is not possible for the opinion with no zealots to be held by more nodes than the opinion with zealots, and this point remains fixed regardless of dimension. In the latter case, the critical point p_c is the threshold value beyond which a stalemate between all k opinions is guaranteed, and the article shows that it decays precisely as a lognormal curve in k. Another study that considers multidimensional opinion model is given in Ref. [101]

1.2 Interaction Networks

In the previous section, we mentioned different types of networks (for example, small-world networks, complete graphs, square lattices, etc.) without defining them mathematically. In this section, we present a summary of relevant notations and definitions, and discuss how networks can be mathematically formalized using matrices. Additionally, we examine key features of social networks and introduce various network models that are used throughout this thesis.

In opinion dynamics models, interaction networks model social influences and play a crucial role in exploring opinion formation. It represents individuals and their connections across a range of social contexts, such as family, friendships, coworkers, and social media. Formally, a network can be described using a mathematical structure known as a *graph*. A graph is composed of a set of N vertices (also known as nodes) denoted by $\mathcal{N}$, and a set of E edges (also known as links) denoted by $\mathcal{E}$, which connect the vertices. The network's edges can be defined using a tuple $(i, j) \in \mathcal{E}$, where i and j are the source and target vertices, respectively. The network can be conveniently represented by an $N \times N$ adjacency matrix A, where each of its elements a_{ij} indicates the connection between vertices i and j, and is defined as

$$a_{ij} = \begin{cases} w_{ij} \in \mathbb{R} & (i,j) \in \mathcal{E} \\ 0, & \text{otherwise}, \end{cases} \tag{1.21}$$

where w_{ij} denotes the weight of the edge between vertices i and j. The out-degree and in-degree of a vertex i are defined as $d_i^o = \sum_j a_{ij}$, and $d_i^i = \sum_j a_{ji}$ respectively. In directed networks, the in-degree and out-degree are not equal. The edge weight w_{ij} is valuable to represent the frequency or intensity of peer interactions, which are not necessarily identical for all pairwise interactions. This also improves the descriptive ability of network analysis, particularly in situations where it is important to differentiate between strong and weak connections between vertices, such as in social influence models. Weighted adjacency matrices are used in various social influence models. For undirected networks, an edge between vertices i and j also implies an edge between vertices j and i, and $w_{ij} = w_{ji}$. Then the adjacency matrix is symmetric, i.e., $a_{ij} = a_{ji}$ and the degree of vertex i is denoted as d_i.

The network can also be represented using a Laplacian matrix L which is an $N \times N$ matrix defined as

$$L_{ij} = \begin{cases} -w_{ij}, & i \neq j, \\ \sum_k w_{ik}, & i = j. \end{cases} \tag{1.22}$$

For undirected networks, the Laplacian can be defined as $L = D - A$, where D is the diagonal degree matrix, with $D_{ii} = \sum_k w_{ik}$.

The network topology, or the way agents are connected within the network plays a vital role in opinion formation. The structure describing the interaction between the agents, along with

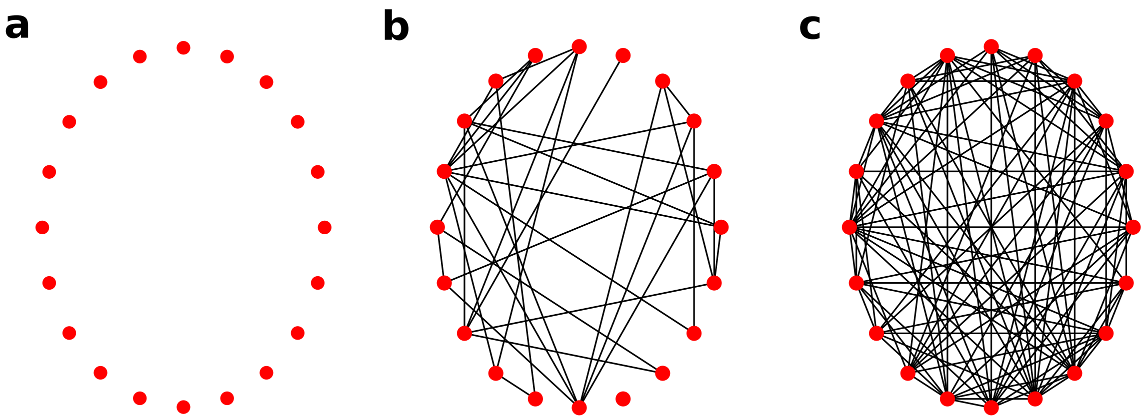


Figure 1.17: The Erdös-Rényi network generated with $N = 20$ and three different probability values of edge connection, **a**: $p = 0$, **b**: $p = 0.2$, **c**: $p = 0.6$.

the frequency and intensity of interaction can significantly impact the spread of information and the formation of opinions. In opinion dynamics models, agents are assumed to be placed at a network's vertices, and the edges define the interaction patterns. There are various types of networks, such as formal, informal, and hybrid networks, with varying topologies. A pyramid-shaped structure represented by a tree is commonly used to illustrate a formal network, which is the prevalent form of organization in sizeable institutions, including corporations, governments, military establishments, and religious bodies [102]. The Erdös-Rényi (ER), scale-free (SF), and small-world networks are some examples of informal networks, which will be discussed in detail later in this section.

Square lattices allow for straightforward representation of opinion configurations and include numerous short loops–a trait shared by real social networks. On square lattices, an agent i assumed to be positioned at the r-th row and c-th column denoted as (r, c) can interact with only four of its neighbors positioned at $[(r-1, c), (r, c-1), (r, c+1), (r+1, c)]$. Initially, opinion dynamics models were frequently examined on a square lattice.

Another simple network structure is *complete graph* in which all agents interact with all other agents in the network. As complex networks have grown in popularity, more research has focused on modeling opinion dynamics on complex social networks, for example, random, scale-free, and small-world networks.

Erdös and Rényi introduced homogeneous random graph in the 1950s, later known as the 'Erdös-Rényi' (ER) network [103]. In the ER model, one starts with N vertices, and an edge is drawn between any two vertices independently with a probability p, $0 \leq p \leq 1$. Fig. 1.17 shows the ER networks generated using three different values of p. The average number of edges in the resulting graph is $m = pN(N-1)/2$ and average degree $k = p(N-1)$. There are different ways of tuning the network parameters to have different topologies. For example, if p is fixed, then $k = p(N-1) \to \infty$, when $N \to \infty$. Alternatively, if $p \to 0$ when $N \to \infty$ in such a manner that $p(N-1)$ remains constant, then k would be finite. Thus, we can obtain graphs of finite and infinite degrees.

The idea of the *Six Degrees of Separation* concept suggests that any two people on Earth are six or fewer connections away from each other. This idea has been incorporated into a widely recognized network model referred to as *small-world* networks. Unlike ER graphs, for

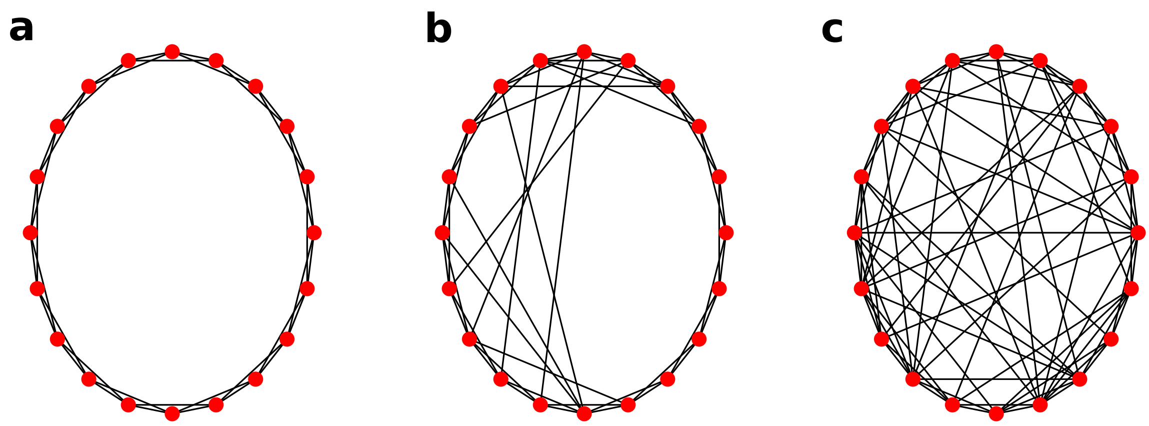


Figure 1.18: The random rewiring procedure for interpolating between a regular ring lattice of $N = 20$ with 4 nearest neighbors and a random network, without altering the number of vertices or edges in the graph following the Watts-Strogatz method with **a**: $p = 0$, **b**: $p = 0.3$ and **c**: $p = 1$

a small-world (SW) network, the average distance between two vertices is small and grows only logarithmically with the network size. Watts and Strogatz [104] introduced a small-world model by introducing a fixed fraction p of random long-distance connections into an initially regular lattice. The construction of the networks is as follows: one starts from a ring lattice with N vertices, each connected to its k nearest neighbors, and each edge is then rewired with a probability p. This architecture enables the change in graph structure between regularity (when $p = 0$) and disorder (when $p = 1$), thereby allowing investigation into the intermediate zone $0 < p < 1$. Fig 1.18 shows the SW network generated using three different values of rewiring probability p.

Watts and Strogatz [104] investigated two primary quantities: the characteristic path length $L(p)$ and the clustering coefficient $C(p)$. $L(p)$ is defined as the number of edges in the shortest path between two vertices averaged across all pairs of vertices. Suppose vertex i has k neighbors; then the maximum number of edges between the neighbors is $k(k-1)/2$ (this occurs when every neighbor of i is connected to every other neighbor). Then $C(p)$ is defined as the fraction of these permitted edges that actually exist in the network, averaged across all vertices. Small-world networks have high $C(p)$ values and low $L(p)$ values.

Since many real networks are dynamic and constantly changing as new vertices join and form connections with older vertices, several models of expanding networks have also been developed. Barabási and Albert [105] proposed a model for a complex heterogeneous network. The construction of the model is as follows: one starts from a complete graph of size m, and new vertices are introduced into the network one by one. Each new vertex selects m other vertices from the existing network by *preferential attachment* rule and creates new edges with them. By the preferential attachment rule, the larger the degree of the vertex is, the more likely it is to get selected. The process terminates when the required network size reaches N. The resulting network has an average degree of $k = 2m$, a clustering coefficient of order $(1/N)$, and probability distribution of the degree of the vertices follows $P(k) \sim k^{-\gamma}$ with $\gamma = 3$. The graphs with power-law degree distributions of $\gamma \leq 3$ are defined as scale-free networks. The network is named the Barabási–Albert (BA) network and Fig. 1.19 shows the 'BA' network

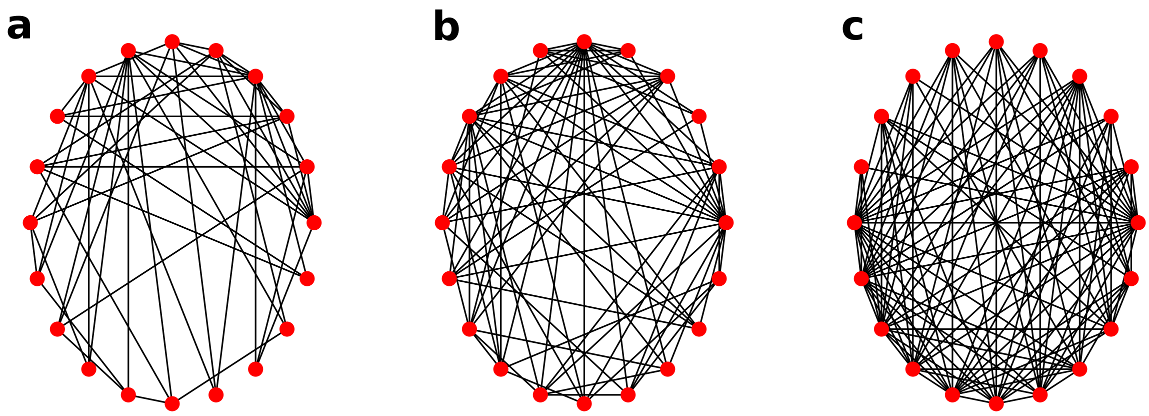


Figure 1.19: The Albert-Barabási network generated for $N = 20$ with three different values of average degree, **a**: $k = 3$, **b**: $k = 5$, **c**: $k = 10$.

generated using three different values of k.

The effect of interaction rules on opinion dynamics models is investigated using different network topologies. Findings suggest that network topology substantially influences opinion consensus [6, 9]. According to the French's formal theory of social power, a population's power structure is tied to the pattern and frequency of interpersonal agreement, which is related to the population's structure of effective communications [10]. In general, it is observed that the system reaches consensus when the agents are well-connected, i.e. similar to a complete graph [10, 31]. For the DW model in scale-free networks, the number of surviving opinions in the steady state is proportional to the number of agents in the interaction network N, for a fixed ϵ [106]. In the DW model, the ϵ_c at which the system reaches consensus is smaller for a complete graph compared to square-lattice, BA and ER networks for $N = 10000$ [31]. Similar results are obtained for the HK model when $N \to \infty$ and $d \to \infty$ (where d is the average degree of the graph). Under these conditions, the consensus threshold is much smaller on a complete graph compared to when d remains finite as $N \to \infty$ [43].

There are many studies on network topologies using directed networks. Gandica et al. [107] studied opinion formation on directed and undirected small-world networks. They concluded that a directed small-world network is more difficult to attain consensus as it has few connecting edges compared to the undirected one. However, network size does not affect the final steady state, although the confidence bound has a more significant influence than small-world rewiring probability [107]. Additionally, there are works on directed and undirected scale-free networks on models for opinion formation [95, 108].

Different network structures are used to determine the probability of a minority winning the election, and showed that certain structures and parameter regimes favor the minority to win [19]. To have more real-world network structures, researchers have generated networks from empirical data using certain assumptions for having an edge between two agents and the corresponding edge weight, and compared the results of discrete opinion dynamics models with the empirical data [109]. The network approach has contributed significantly to understanding various complex systems' structure, function, and response [9].

1.3 Validation

Opinion dynamics models must undergo validation to verify whether they accurately reflect the occurrences or behavior the models are supposed to depict in real life [110, 111]. Increasing the model's accuracy entails comparing the model's predictions against actual observations or experimental data. To effectively capture real-world patterns, it is essential to consider the social interaction rules employed by agents, their interaction network configurations, the distribution of their initial opinions and intrinsic preferences, and the relative importance of social information. Empirical measurement may be used to constraint the model parameters, and the resulting models can be used to study opinion formation in real societies [109].

Validation of a model should employ empirical data, and collecting them could be challenging. We need to put something as complex and multifaceted as an opinion into a quantitative metric that can be estimated from the available information. Even when opinions are conformed to a few options, as in sociological surveys and questionnaires, data collection efforts are constrained by sample size and answer subjectivity [112]. Additionally, these surveys cannot guarantee the extent of honesty of the agents.

Elections and the surveys that precede them are notable sources of empirical data, as they provide population-level statistics on political candidates of choice as well as the motivations and attitudes that drive voters' actions [110]. Another source of empirical data is from social media platforms. We can assess the agreement of a user towards an issue or a political party by extracting their comments/posts in social media [111]. Nonetheless, due to the scarcity of data, modeling and simulation have become the dominant methods in the field. In these methods, mechanisms derived from sociological theory (or otherwise deemed reasonable) are converted into mathematical rules and analyzed, frequently referring to common sense rather than empirical validation [6, 113].

Opinion dynamics models can be validated with empirical data in various ways. The study of opinion dynamics models, if properly validated, offers information on the social processes behind knowledge transfer, public opinion, and group cohesiveness. The model's adaptability allows us to look into a variety of effects such as the impact of social media (e.g., an external field influencing all agents) [8, 114], the impact of technology [4] and algorithmic bias to the adaptive interactions between opinion dynamics and social network structure [111].

Filho et al. [78] performed a statistical analysis of the Brazilian election of the year 1998. Let $N(v)$ denote the number of candidates securing a fraction v of the total votes. They showed that $N(v)$ follows a power law distribution, i.e., $N(v) \propto v^{-\alpha}$ with $\alpha \approx 1$. Further studies employed Sznajd model and confirmed the result from Ref. [78] using square lattice, cubic lattice and Barabási network [82, 115]. A further examination on Indian elections revealed a similar pattern, implying that $N(v)$ may reflect some degree of universality [116]. The results from the same study extended to German, French, Italian, and Polish elections showed a significant deviation from the power-law of $1/v$ for different countries, proving that the pattern is not universal to all countries and eras [117].

Later on, additional datasets of a number of nations and electoral periods showed a lognormal distribution for $N(v)$. Let v_0 denote the average number of votes cast for all candidates in the same party. In order to combine the numerous curves of $N(v)$ vs. v/v_0 into a single universal (lognormal) function, v_0 must be included as a scaling factor and $N(v)$ is given by

$$N(v) = \frac{v_0}{\sqrt{2\pi}\sigma v} e^{-(\log(v/v_0)-\mu)^2/2\sigma^2}, \tag{1.23}$$

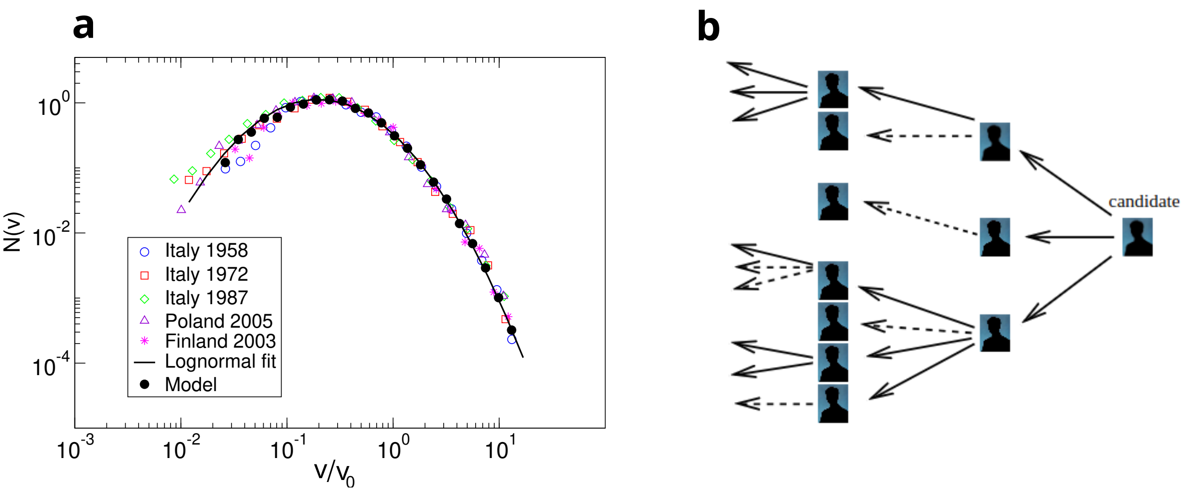


Figure 1.20: **a**: Universality of the scaling function $N(v)$ across different countries and years. The lognormal fit, performed on the Polish curve, describes very well the data. The universal curve is well reproduced the dynamical model, where the dynamics of the voters' opinions reflects the spreading of the word of mouth in the party's electorate. **b**: Spreading of the word of mouth among voters. The candidate (right) convinces some of his/her contacts to vote for him/her. The convinced voters become "activists" and try to convince some of their acquaintances, and so on. Successful interactions are indicated by solid lines, unsuccessful interactions are displayed as dashed lines.*Figure and caption taken from Ref. [117]*

with $\mu = -0.54$, $\sigma^2 = -2\mu = 1.08$. These election outcomes are analyzed using a straightforward opinion dynamics model based on word-of-mouth. Candidates are the nodes of a tree-like network of voter interactions, and the number of contacts follows a power-law distribution. Voters who have already decided about a candidate try to persuade their peers to support the same candidate (see Fig. 1.20 **b**). Initially only candidates have an opinion (they cast vote for themselves). The candidates begin the dynamics by attempting to persuade their neighbors. Each candidate's supporters develop into activists who then attempt to get their neighbors to support them, and so on. Only undecided voters can be convinced by persuasion. Not all interactions convince an undecided voter: persuasion occurs only with probability p. The fit of election data with the model is shown in Fig.1.20 **a**.

Galesic et al. [109] compared the predictions of discrete models with $x_i = \pm 1$ opinions to real-world survey data. The satisfaction of an agent regarding their particular opinion is then defined as

$$S_i = \alpha x_i h_i^{\text{soc}} - (1-\alpha) x_i h_i, \tag{1.24}$$

where $0 \leq \alpha \leq 1$ denote the relative weight between the influence of neighbors opinion and their individual preferences, h_i is the intrinsic preference to particular opinion, h_i^{soc} can take different values depending on the model,

$$h_i^{\text{soc}} = \begin{cases} x_j, & \text{for voter model,} \\ sgn(\sum_{j=1}^{g} x_j), & \text{for majority model,} \\ x_e, & \text{for presence of zealots.} \end{cases} \tag{1.25}$$

The authors used the survey data on 80 individuals living in MIT dorm during the 2008 presidential election period and another survey data on 94 individuals recruited from Mechanical Turk during the 2016 presidential election season. The study was conducted on complete graph, ring lattice, small-world network, *objective empirical* (the proximities of Bluetooth signal of two agents' cell phones determine the edge and the weight of the edge by the likelihood that the two agents are on the same floor) and *subjective empirical* (edges are determined by collecting sociometric data from agents about their contacts and weighted using the frequency of interaction) networks. They showed that the models can be parameterized to resemble actual societies, and can reproduce opinion dynamics on agent and group levels .

There are a few other studies that include data of historical elections. In Ref. [118], the election results have been analyzed using Wikipedia data while modeling electoral processes. The data on vote-share distributions at the county level in US presidential elections from 1980 to 2012 is replicated by the voter model [119]. The noisy voter model explains data on the Lithuanian parliamentary elections of 1992, 2008, and 2012, which generates a beta distribution of vote shares [120, 121]. To uncover underlying social impact mechanisms, Phan et al. [122] fitted a binary-opinion model to the poll data from the US presidential elections from 2012 to 2016.

1.4 External Influence

Information can spread through various medium such as word of mouth, newspapers, radio, mass media, and in modern era via Internet, and social media. Most of the models we have examined so far apply to frameworks where peer-to-peer interactions attempt to bring about consensus among populations. However, we are currently inundated with an enormous amount of external influences - here, 'external' denotes that this influence originates from sources other than peer interactions, for example, social media.

The impact of mass media is investigated using the Sznajd model with the agents placed on the vertices of a square lattice [123]. The model involves the introduction of an external influence, the media, which has a specific value (say $+1$). If four neighbors agree on a particular opinion, then all their other neighbors switch to the former's opinion. If they disagree, the media's opinion is adopted with a probability of p. The research demonstrated that the final state (either ± 1) depends on the initial density of $+1$ opinions and the value of p. A higher value of p requires a lower initial density of $+1$ for complete agreement with the media. When $p \gtrsim 0.18$, the population always converges to the value of the information from the media. Similar work with three opinion states is performed in Ref. [124].

González-Avella et al. [125] investigated the impact of external influence on a binary model that can take two distinct opinions. At different time intervals, external influence could choose to hold one opinion with a probability $p > 0.5$ (the true or most beneficial opinion) or the other value with a probability $1 - p$. A random agent is selected to interact with this external influence at each time step. If the agent had a different opinion, it would only be updated if a fraction of its neighbors, larger than a threshold value τ, shared the same opinion as that of the external influence. The study demonstrated that the system could reach consensus only for intermediate values of τ. Mixed populations with fluctuations were obtained for small values of τ, while for large values of τ, the population remained in its initial condition.

Laguna et al. [126] investigated non-equivalent binary opinions in which one of them is considered as the right one. Agents opinions are updated based on small group interactions in which a poll determines whether or not to alter the opinion. In this poll, the higher value of one opinion counts, and a weight is used for the self opinion. Instead of communicating with peers, agents communicates with probability p with a so called 'monitor', which forces them to accept the right opinion. This method introduces external influence where the external field's persuasion is infinite. A different approach would be to use a group of zealots. The two options have been found to enhance the adoption of the correct option, however the introduction of zealots was less effective than the monitors.

The effects of external influence have been investigated in continuous opinion models as well. The dynamics of DW model is used to investigate the effects of external influence [127]. Every agent in the population is subjected to an external influence x_e that promotes a specific opinion. At every T time step, the whole population interacts with this external influence, with the same condition of interaction. An agent's opinion changes if the difference between the opinion of agents and the external influence is less than the confidence bound. Experiments were carried out using $\mu = 0.5$. The system dynamics depend on the value of external influence, T, and the confidence bound parameter ϵ. The population converges to the value of external influence if the confidence bound is large enough to reach all agents. Conversely, reaching a consensus with the value of external influence is impossible if the confidence bound is too small.

A modification of the DW model is used to analyze the effects of external influence (mass media) where after certain number of updates, all agents interacts with external influence x_e controlled by a parameter γ [128]. Then the system dynamics evolve as,

$$x_i = x_i + \mu\gamma\epsilon_i(x_e - x_i) \tag{1.26}$$

where $\epsilon_i = 1 - \phi|x_i|$ is the confidence bound of agent i and ϕ controls the tolerance rate of agents. Agent opinions move towards the value of x_e for mild information (low γ), while an increasing number of antagonistic clusters arise for strong information. This demonstrates that aggressive media campaigns are dangerous, as they may result in the population not receiving information.

A different model similar to the DW model investigated the effect of external influence and disagreement [129]. Disagreement is introduced into the model as an attribute $w_{ij} = \pm 1$ of the link between two agents (some pairs of agents always agree and some others always disagree), and opinions of agents $x_i \in [0, 1]$. Then the dynamics of the system evolve as,

$$x_i(t+1) = x_i(t) + \mu\omega_{ij}(x_j(t) - x_i(t)) \tag{1.27}$$

where $\mu = 0.5$ is the convergence parameter. The media impact is parametrized as $x_e \in [0, 1]$ in the model, which can simultaneously alter the opinions of all agents in the network after a certain number of updates. The introduction of disagreement among some pairs of agents was shown to favour consensus with the external influences. Another study using DW model included the competition among multiple interacting mass media (for example, TV, blogs, newspaper, social media platforms, and so forth) sources [130]. The authors concluded that media competition tends to favor fragmentation among the population.

Quattrociocchi et al. [131] investigated a two-dimensional DW-like model with conflicting opinions (x_i^1, x_i^2) on the topics of welfare and security. The study focused on the impact of external influence from media and experts on a scale-free social network. The results showed that peer interaction could help agents escape false media messages, but only if the media does not reach more than 60% of the total population.

Hegselmann and Krause incorporate the truth seekers, i.e., agents that take into account the value of truth $T \in [0,1]$, into the Hegselmann-Krause model [132]. This can be considered as agents interacting with experts, which is analogous to interacting with an external influence. As an agent interacts with their neighbors, they readjust their opinion following,

$$x_i(t+1) = \alpha_i T + (1-\alpha_i)\frac{1}{|N_i(t)|}\sum_{j \in N_i(t)} x_j(t), \tag{1.28}$$

where $N_i(t)$ denotes the neighbors of the agent i at time t, and $\alpha_i \in [0,1]$ depicts the agent's disposition to seek the truth. The effect of the truth T is not based on confidence bound; rather, it affects agents with $\alpha_i \neq 0$ independent of their opinions. Given that the truth is not extreme (near to 1), the results shows that even for small $\alpha_i = 0.1$ for all agents, or if at least half of the population $\alpha_i \neq 0$, the population converges to the truth value. If the truth is extreme and not all agents have $\alpha_i \neq 0$, some agents remain far from the truth. A large α_i value may cause more agents with $\alpha_i = 0$ to stay away from the truth, implying that too strong information can have reverse effect. Kurz [133] presents a more in-depth examination of the model with truth seekers, demonstrating analytically that all truth seekers (agents with $\alpha_i \neq 0$) converge to the truth. This result holds even if there are agents who do not seek the truth.

The effect of an external influence on the Axelrod model is extensively studied by introducing external agent (information source, information field) that can interact with agents in the population. One approach is to include a parameter p that specifies the probability that an agent will engage with the external influence instead of a neighboring agents at each time step [134–137]. Increasing the probability of interaction with external influences increases fragmentation instead of favouring consensus. This behavior is explained by the fact that interacting with an external influence more frequently decreases agents' probability of interacting with one another. As a result, relationships between agents and external influence are interdependent. In addition, the fact that some agents cannot interact with the external influence due to their low overlap (defined by Eq. (1.20)) with the agents causes an isolation of these agents and results in the formation of new clusters.

The external influence in the above mentioned papers was constant and independent of the population state. González-Avella et al. [138, 139] investigated several other ways to define external influence (so called global and local endogenous fields), on a two-dimensional lattice. They accounted for endogenous cultural influences on Axelrod model by calculating the statistical mode of opinions over the entire population (global) or each agent's neighborhood (local). The results are similar to the previous studies, where these fields encourage population fragmentation for a large value of p; however, for small values of p, the population converges to the external influence. Local external influence sources encouraged population to reach consensus while revealing quantitative discrepancies between the different types of information. Furthermore, an analysis of two distinct populations, where each population is influenced by a global field of the other, has revealed complex behavior, where sometimes one population did align with the influence from the others, but they could also reject it or form a large rejecting minority [138]. Several studies were conducted to develop techniques to tackle the interdependence between external influence and agents interactions [140–142], and their extensions to combine the effect of media with noise and social network structure [142, 143].

Sîrbu et al. [144] introduced an approach for modeling continuous vectorial opinions in a complete graph, which was later extended to different topologies, including external influence and disagreement. In this model, opinions are represented by a point in the simplex in $(K-1)$-

dimensions, $\boldsymbol{x} = [x_1, x_2, \cdots, x_K]$. The interactions among the agents are defined based on a similarity measure defined as

$$o^{ij} = \frac{\sum_{k=1}^{K} x_k^i x_k^j}{\sqrt{\sum_{k=1}^{K} (x_k^i)^2 \sum_{k=1}^{K} (x_k^j)^2}}, \tag{1.29}$$

determining the probability of agreement or disagreement (opinions become more similar or dissimilar) between two agents. External influence is introduced into the system as a static agent. Following agents interaction, each agent can interact with the static agent with probability p. The system was demonstrated to generate one or more clusters based on the value of p, initial condition, and external influence. The study showed that extreme external influence and large exposure had a lower success rate in the population, whereas mild external influence and low exposure were more easily accepted by many agents. Full agreement with the external influence was achieved only for a very mild external influence or a very low exposure of a non-extreme external influence to an initial configuration. Sîrbu et al. [145] further developed the model, where multiple external influence (for example, multiple mass-media) sources were analyzed. It is shown that the system leads to stable non-polarized clusters or full agreement for external influence which is not extremely mild.

Most analyses have examined only one static source of information, whereas, in reality, information is derived from multiple sources and is continually evolving [8]. These topics have received little attention in the literature, and challenges persist in establishing a framework where media and agents interact bidirectionally, similar to society.

1.5 Electoral Systems and Computational Propaganda

In modern democratic societies, elections are typically held periodically and involve various activities including campaigning, casting votes, counting ballots and finally translating into elected representatives. Elections may be held at the local, regional, or national level and can involve a variety of political parties and candidates. Finally, the outcome depends on how the individual votes are gathered and counted, which is different in different countries and branches of government.

Electoral attacks refer to any attempt to disrupt or manipulate the electoral process. These attacks can take many forms including cyber attacks on electoral systems, disinformation campaigns to spread false information about candidates, and physical attacks on voters or polling stations. There are several recent reports of systematic attempts to influence the outcomes of democratic elections both internally [146] and externally [20]. Studies reveal that social media have been used widely for electoral attacks in various countries using computational propaganda especially through bots, and thus having a negative impact on the elections [147].

Electoral attacks can have serious consequences for the democratic process, as they can undermine the integrity and legitimacy of elections. To counter electoral attacks, governments and electoral authorities need to take steps to ensure the security and integrity of the electoral process. This may involve educating citizens about the risks of electoral attacks and how to spot disinformation campaigns can help to promote awareness and resilience in the face of these threats. It is essential to impose rules on social media platforms to differentiate the information created by the bots from humans. Ultimately, protecting the integrity of the electoral process is essential to ensure that citizens have confidence in their government and can participate freely and fairly in the democratic process.

The advent of social media has revolutionized the way people communicate, interact, and share information. With its rapid proliferation, social media has also profoundly impacted various aspects of society, including the political landscape. One of the most significant arenas in which this impact is felt is in the realm of elections. The influence of social media on elections is multifaceted, ranging from its ability to shape public opinion and mobilize voters to its potential for spreading misinformation and manipulating political discourse.

While social media has the potential to enhance political participation, it also harbors the danger of spreading misinformation and disinformation. False narratives, misleading content, and fake news can spread rapidly across these platforms, influencing voters' perceptions and decisions. The speed at which information travels online often outpaces fact-checking efforts, leading to the persistence of inaccurate information. The use of bots–software-created and operated, automated social media accounts– for spreading disinformation during elections can have significant and concerning impacts on the democratic process and public discourse. Bots can amplify false or misleading information, creating the illusion of widespread support for a particular viewpoint or candidate. Bots can flood social media platforms with automated messages, making it difficult for genuine conversations to take place. This can sway public opinion and influence voter decisions based on inaccurate or biased information. This darker side to social networks and their use in democratic processes goes by the name *computational propaganda* [18, 148–150].

There are several examples of recent massive uses of bots during political campaigns. Bots appear to have been extensively employed to influence perceptions of the UK's role in the European Union during the 2016 Brexit campaign [151]. During the 2016 United States presidential election, social media platforms were utilized to disseminate significant volumes of false information [150, 152]. In the context of the 2017 United Kingdom general election, the labor party employed bots to disseminate electoral messages via Tinder [150]. Similarly, in the 2018 Brazilian presidential election, the incumbent candidate Jair Bolsonaro extensively employed social media channels to disseminate targeted information [153]. The 2019 Indian general election saw both the ruling Bharatiya Janata Party and the Indian National Congress making widespread use of online misinformation as part of their campaign strategies [147].

Keeping computational propaganda in check requires a combination of technological, regulatory, and educational measures. This can include anti-bot measures such as CAPTCHAs, laws and regulations that address the misuse of automated systems for propaganda purposes, promote digital literacy and educate users about the techniques and strategies used in computational propaganda, etc. In this thesis, we look at the problem from another novel perspective and investigate: (i) what social characteristics of a population of voters may render it more resilient against external efforts to manipulate its opinion and (ii) which of the existing democratic electoral systems is more robust against computational propaganda. These important questions have not been addressed before, to the best of our knowledge.

In order to address these questions, it is important to understand the different types of democratic electoral systems that exist around the world. Democracies are political systems with different governmental branches. Different democratic countries have in general different electoral processes, which also differ in elections to different branches of government – executive, legislative and judiciary [154]. Roughly speaking, we can group electoral systems into three main categories, which are single-winner, multi-winner, and mixed systems. Fig.1.21 shows the electoral systems practised in their corresponding country, and some countries are not democratic, for example, China [155, 156].

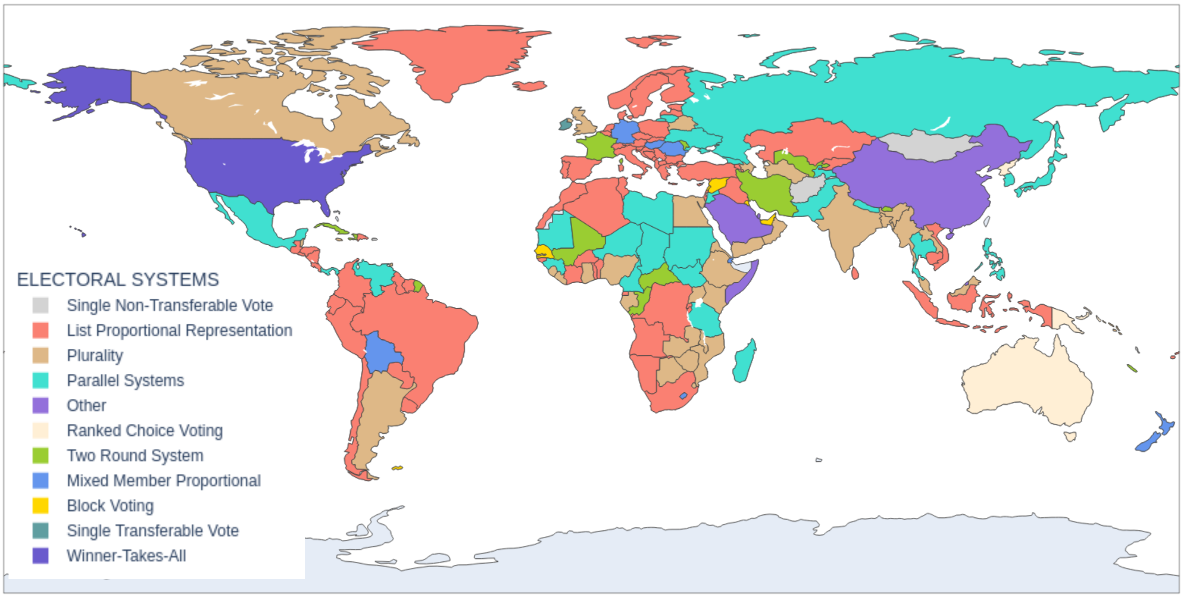


Figure 1.21: Countries with different electoral systems.

- Single-winner system : It can be sub-classified into three different sub-systems.
  - Plurality: The system in which the candidate with the most votes wins, without necessarily a majority of 50% votes (for example, Canada).
  - Two-Round System: This system is similar to plurality. If no candidate attains a majority of more than 50%, the second round of the election is conducted between the two candidates who won most votes in the first round of the election (for example, France).
  - Ranked Choice Voting (single winner): The electors are given a choice to rank the candidates. If the first-priority votes cannot conclusively help decide the winner, then the second-priority votes of candidate with least number of votes are distributed among the other candidates. This process continues until one candidate wins the majority of the votes (for example, Australia).
- Multi-winner system: This is commonly used when a country is divided into electoral districts, and representatives have to be selected to fill m seats from each electoral district. When selecting m representatives for a legislative body, the seats are allotted in proportion to the number of votes received or all the seats are allotted to the party with most of the votes. This can be achieved in different ways.
  - Block Voting: A system in which the electors have as many votes as the number of representatives to be elected. The number of votes is counted as per the plurality system. Electors are unable to vote more than once for the same candidate. The m candidates with the most votes are declared elected and will fill the seats (for example, Syria).
  - Single Voting (Single Non-Transferable Vote): A system in which the electors have just one vote. The candidate with the most votes becomes the representative (for example, Afghanistan).

- List Proportional Representation: A system in which parties nominate their candidates and the electors vote for the candidates nominated by the parties. The number of seats allotted to each party depends on the proportion of votes for that party (for example, Switzerland).
- Winner-Takes-All: A system in which all seats are allotted to the party with the most votes even if only a small part of the majority supports the party (for example, United States).
- Cumulative Voting: A system in which electors have the number of votes equal to the number of representatives to be elected. Electors can assign votes according to their wish. These votes can be distributed among candidates or assigned to a single candidate (for example, Canada).
- Proportional Ranked Choice Voting (Single Transferable Vote): A method of election in which voters have only one vote but are given a choice to rank the candidate as per their preference. The m candidates with the most votes are declared elected and will fill the seats (for example, Ireland).

- Mixed Systems: A system which combines single winner and multi-winner proportional systems. These can be divided into two types.
  - Mixed Member Proportional: A system in which electors get two votes, one for the candidate and another for the party. One fraction of the seats is elected using the plurality system and the other uses list-proportional systems. The list seats are assigned after selecting plurality seats such that they are proportionate to the national party vote (for example Germany).
  - Parallel systems: A system in which electors get two votes, one for the candidate and another for the party. One fraction of the seats are elected using the plurality system, and the other uses list proportional systems. The list of seats was allocated proportionally to the national party vote (for example, Nepal).

Below, we develop an opinion dynamics model, where a voter's opinion evolves depending on their *natural opinion* and their interactions with other voters. Here, the natural opinion refers to an agent's opinion in the absence of interaction and external influence. With the addition of interaction, the natural opinions evolve to reach an equilibrium known as the final opinion. These final opinions are compiled following the electoral system used, to generate the final outcome of the election. We then study the effect of external influence by sequentially adding a small bias to the initial opinion of each of the agents, one by one. We quantify the measure of robustness of an election process as the minimum number of agents required to be influenced (which we call the *effort*) in order to overturn the election result.

We study the impact of the social characteristics of a population on election results by systematically varying the distributions from which we draw the initial opinions. Whereas, we investigate the robustness of different electoral systems by examining the 'effort' when using them for forming the election results. We remark that our models must be used only to understand some of the characteristics of election processes at a qualitative level. One main result in the thesis is that, we validate our simulation results with historical election results of the US House of Representatives from 2012 to 2020. We observe a significant correlation between the simulation data and actual election data, which we consider quite remarkable.

Our model builds upon and extends the current voter and opinion dynamics models. It includes several aspects that were previously considered only separately. For instance, it considers the finite interaction distance between agents, termed the confidence bound in the Deffuant [29] and Hegselmann-Krause [15] models. Additionally, it takes into account the natural opinions that voters would hold if they were not influenced by others. This particular feature helps avoid artificial aspects like consensus and fragmentation seen in opinion dynamics models with limited confidence bounds [157]. Moreover, this feature enables the modeling of different populations with varying levels of opinion polarization, allowing us to explore how this polarization affects electoral processes and their resilience. A key innovation in our model is its ability to extend consensus-like dynamics to multi-dimensional spaces where each dimension corresponds to a different political party. This involves incorporating party ranking from left to right without favoring any particular side, which our model does explicitly. Lastly, we distinguish ourselves from standard voter and opinion dynamics models by introducing outcome functions designed to capture essential characteristics of diverse electoral processes seen in different democratic countries for various branches of government.

The mathematical details and the findings of the defined problem will be discussed in the following chapters. It's important to note that we won't cover all the mentioned electoral systems, but rather focus on the main ones that are, (i) Single representatives, (ii) Winner-takes-all representatives, (iii) Proportional representatives and (iv) Ranked choice voting system. They encompass around 80% of the electoral systems worldwide.

1.6 Appendix

1.6.1 Linear Algebra Concepts

Here, we recall some of the essential linear algebra definitions and results necessary for our discussions.

Definition 1.6.1. *A square matrix M is diagonal dominant if for all i,*

$$|M_{ii}| \geq \sum_{j \neq i} |M_{ij}|. \tag{1.30}$$

Theorem 1 (Gershgorin's Circle Theorem)**.** *Let M be a square matrix with elements $M_{ij} \in \mathbb{C}$ and define the Gershgorin discs,*

$$D_i := \Big\{ z \in \mathbb{C} \big| |z - M_{ii}| \leq \sum_{j \neq i} |M_{ij}| \Big\}. \tag{1.31}$$

Then each eigenvalue of M lies in the union of the Gershgorin discs, $\bigcup_i D_i$.

Corollary 2. *Let M be a diagonal dominant matrix with positive (resp. negative) diagonal elements. Then all eigenvalues of M have non-negative (resp. non-positive) real part.*

Definition 1.6.2. *An $n \times n$ symmetric real matrix M is said to be positive-definite matrix if $x^T M x > 0$ for all non-zero $x \in \mathbb{R}^n$.*

Definition 1.6.3. *A real symmetric positive definite matrix with non-positive off-diagonal entries is a Stieltjes matrix.*

Theorem 3 (Perron–Frobenius Theorem)**.** *Let $M \in \mathbb{R}^{n \times n}$ be irreducible and non-negative matrix, and suppose that $n \geq 2$. Then,*

(a) $\rho(M) > 0$

(b) $\rho(M)$ *is an algebraically simple eigenvalue of* M.

(c) there is a unique real vector $\boldsymbol{x}$ such that $M\boldsymbol{x} = \rho(M)\boldsymbol{x}$ and $x_1 + \cdots + x_n = 1$; this vector is positive.

(d) there is a unique real vector $\boldsymbol{y}$ such that $\boldsymbol{y}^T M = \rho(M)\boldsymbol{y}^T$ and $x_1 y_1 + \cdots + x_n y_n = 1$; this vector is positive

2 Two-party System

We start with a simple case of two-party system, with an electoral framework of two dominant political parties or the case of 'yes/no' decisions. Examples of two-party systems include elections in United States, referendums etc. A two-party electoral system has a straightforward structure: two main parties compete for voters' favor, often leading to a clear-cut outcome. Rooted in a binary competition, characteristics of a two-party system profoundly impact representation, governance, and political discourse by shaping the choices available to voters, influencing policy positions, and affecting the overall political climate.

The outcome of an election is the result of a complex process. Indeed, the election outcome depends on the opinion of each individual in the voting population, but there are other factors as well. For example, it depends on the interaction between individuals and also on how the individual votes are gathered and counted. We model an electoral process using an opinion dynamics model which considers opinion of the individuals, interaction between them and some external influences that tries to modify individual's opinions. We assume that each individual in the model votes in favor of the party that aligns most closely with their opinion. The votes of the individuals are aggregated at different levels of governance and electoral system to determine the election outcome corresponding to a priori polarization and bias. The detailed description of the opinion formation model that we consider and formalization of various electoral systems that we investigate is provided later in this chapter.

Electoral systems are often subjected to different types of external attack in attempts to change the election outcome. These external attacks can have serious consequences such as inducing polarization in the society. In order to maintain the well-functioning of a democratic system, it's essential that the electoral system by which the parties or candidates gain power be robust to external attacks. There have been recent report of attempts to influence the outcomes of democratic elections externally. With different electoral systems being used among democratic countries, a question that naturally arises is whether certain electoral systems are more robust against external influences than others. Our goal is to compare the relative robustness of different electoral systems against systematic external efforts that attempt to modify agents opinions.

2.1 Opinion Dynamics

In our model, voters are represented as agents. Each agent has an opinion, that may change upon interactions with other agents. We model an agent's opinion as a real number that indicates the agent's degree of adhesion to a party. In two-party systems, we define positive opinion values as the inclination towards one of the parties and the negative values to the other. Let x_i denote the opinion of agent i. The sign of the agent's opinion indicates which party agent i supports, and its magnitude represents the degree of agreement or support for that party.

In the existing opinion dynamics models, agents generally reconsider their opinion based on interactions with other agents. As time evolves, they forget their own opinion to reach a common consensus, which is unrealistic in the real world. As, it is obviously uncommon for many different agents to have exactly the same opinion on a given, complex topic. We adapt our model to prevent this unrealistic behavior by introducing a *natural opinion* $x_i^0 \in [-1, 1]$ to each agent in the population. The natural opinions of an agent is the actual opinion they would have in the absence of agent-agent interactions.

The agents in the population tend to interact more with other agents of similar opinions than those with different opinions. We consider that two agents within an electoral unit (the region in which election is conducted) interact if their natural opinions are sufficiently close, i.e., below a confidence bound threshold $\epsilon > 0$, and agents belonging to two different electoral units will not interact. This confidence bound ϵ determines the interaction network in our model. The elements of the adjacency matrix A_ϵ^u of the interaction network for an electoral unit u are given by

$$(A_\epsilon^u)_{ij} = a_{ij} = \begin{cases} 1\,, & \text{if } |(x_i^0)^u - (x_j^0)^u| < \epsilon\,, \\ 0\,, & \text{otherwise.} \end{cases} \tag{2.1}$$

The corresponding Laplacian matrix is defined as

$$L_\epsilon^u = D_\epsilon^u - A_\epsilon^u\,, \tag{2.2}$$

where $D_\epsilon^u = \text{diag}(d_1, \cdots, d_n)$ is the diagonal degree matrix of the interaction network. The interaction network remains constant throughout the dynamics (i.e. matrix A_ϵ^u is time-independent) and is determined only by the natural opinions of agents.

The dynamics of the model is given by

$$\dot{x}_i = d_i^{-1} \sum_j a_{ij}(x_j - x_i) - (x_i - (x_i^0 + \omega_i))\,, \tag{2.3}$$

with ω_i being the external influence applied on agent i (see Sec. 2.1.2). We normalize the total influence of other agents on agent i using its degree (number of neighbors) d_i. This implies that, the influence that an agent has on a given agent i is diluted by the total number of agents interacting with agent i. In vector form, for an electoral unit u with n_u agents, Eq. (2.3) is rewritten as,

$$\dot{\boldsymbol{x}}^u = (\boldsymbol{x}^0)^u + \boldsymbol{\omega}^u - \left[(D_\epsilon^u)^{-1} L_\epsilon^u + \mathbb{I}\right] \boldsymbol{x}^u\,, \tag{2.4}$$

where $\mathbb{I}$ is the n_u-dimensional identity matrix.

Eq. (2.4) is a linear system and the matrix $(D_\epsilon^u)^{-1} L_\epsilon^u + \mathbb{I}$ has strictly positive eigenvalues (Lemma 4 in Appendix 2.7). Therefore, the dynamics is globally stable and converges exponentially fast towards the equilibrium

$$\boldsymbol{x}_\infty^u = \left[(D_\epsilon^u)^{-1} L_\epsilon^u + \mathbb{I}\right]^{-1} \left((\boldsymbol{x}^0)^u + \boldsymbol{\omega}^u\right)\,, \tag{2.5}$$

which gives the final opinions of agents.

2.1.1 Natural Opinions

Electoral populations are defined by two main social characteristics which are their openness of the society (modeled by the confidence bound ϵ), and the distribution of the agent's natural opinion. The natural opinion of the agents are modeled based on their bias towards one party, their polarization and their width of represented opinions. Following Refs. [158, 159], we consider bigaussian probability distributions given by

$$P(x) = \frac{p}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu+\frac{\Delta}{2})^2}{2\sigma^2}} + \frac{1-p}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu-\frac{\Delta}{2})^2}{2\sigma^2}}, \tag{2.6}$$

where $x \in [-1, 1]$, $\sigma > 0$ is the standard deviation of each Gaussian peak. The $\Delta \geq 0$ quantifies the degree of polarization, $\mu \in \mathbb{R}$ global shift of the population, and $p \in [0, 1]$ gives the weight of gaussians. In this thesis, the width of the distributions is fixed at $\sigma = 0.2$, without loss of generality. The total mean of the bigaussian probability distribution $P(x)$ is calculated as

$$\begin{aligned} \mu_{\text{total}} &= \int_{-\infty}^{\infty} xP(x)dx \\ &= p(\mu - \frac{\Delta}{2}) + (1-p)(\mu + \frac{\Delta}{2}). \end{aligned}$$

In the absence of any bias towards one of the parties, the distribution of natural opinions can be unpolarized or polarized. The value of polarization parameter Δ ranges from $\Delta = 0$ (an unpolarized population, whose natural opinions follows a simple Gaussian), to $\Delta > 4\sigma$ (polarized population, whose distribution of natural opinion gives two well distinct Gaussian peaks).

There are two different ways of introducing bias in a society, either (i) via differences in weight p between the two Gaussians, or (ii) by shifting μ towards one of the parties. In our studies, we change the percentage of agents voting for each party by tuning these two bias parameters.

2.1.2 External Influence

The vector of external influence $\boldsymbol{\omega}^u \in \mathbb{R}^{n_u}$, indexed by the electoral unit u, models the targeted influence of an external source on an agent. It can represent the impact of a wide range of sources from mass media to targeted advertisement through online social media or computational propaganda. Without loss of generality, we set the '+1' party as winner in the absence of external influence. The external influence aiding the opponent party tries to change the outcome by stirring a set of agents towards the '−1' party, i.e., we have $\omega_i^u \leq 0$, for all i.

To be successful, influences requires to engage a certain amount of resources. In our simulations, we proceed iteratively to determine the effort required to swing the result of an election at the level of an electoral unit. Starting from an uninfluenced system ($\boldsymbol{\omega}^u = 0$), we add a constant small negative value $-\delta$ to a components of $\boldsymbol{\omega}^u$, repeat to other components, until the outcome of the election changes in the electoral unit. In our numerical simulations, we set $\delta = 0.1$. If the outcome of the election in electoral unit u changes after K_u increments (representing the number of agents influenced) , we define the *effort* needed to change the election outcome as the percentage of agents influenced,

$$\xi_u = \frac{K_u}{n_u} * 100\,. \tag{2.7}$$

Next we repeat the process in other electoral units, until the election results is changed at the country level. The total effort needed to change the outcome of the election is

$$\xi = \sum_u \frac{n_u}{N} \xi_u \,, \tag{2.8}$$

where N is the total number of agents in the whole country.

When influencing agents in order to change the election outcome in an electoral unit, we define strategies of influence to obtain the maximum impact with minimal effort. Intuitively, one would target the agents close to the center of opinion distribution in priority as their vote is likely to be easier to change. We confirmed that this strategy is efficient by comparing the effort needed to change the outcome by influencing the agents closer to the center of the opinion space to the effort when target agents at random (see details in Appendix 2.7.1).

When a country is composed of multiple electoral units, there are two natural ways to select which electoral unit to target in priority. One can either target agents in the electoral unit with the lowest population, because changing the outcome in such a unit requires to influence less agents, or one can target the electoral unit with smallest relative majority, because such a unit is close to a change of outcome already. We show that targeting the electoral units with lower relative majority is generally more efficient than targeting the electoral units with lower population (see Appendix 2.7.2). Our strategy to overturn elections is therefore to target weakly opinionated agents in electoral units with lower relative majority.

2.2 Electoral Outcomes

Once the final opinion of each agent is known, the election outcome depends on the levels of vote aggregation and the type of electoral systems considered. The sign associated to the opinion of each agent represents the party they vote for, i.e., we refer to the two parties as '+1' and '−1'. We assume that each agent votes for the party that is the closest to their opinion, i.e., agent i in the electoral unit u votes for party $\text{sign}(x^u_{\infty,i}) = \pm 1$.

2.2.1 Electoral Units

Elections are conducted at various levels, depending on the specific country and the governmental body for which the election is being held. Here we define three *electoral units* corresponding to three different levels of vote aggregation.

Country: In almost all electoral instances, the largest possible electoral unit is a country. Presidential elections are often realized at the level of country, with notable exception of parliamentary elections in the European Union. Examples of elections realized at the country level include presidential elections in Chile, France, Ireland, Mexico and several other countries.

State: Democratic countries are partitioned into states typically with different populations. Other denominations for states include provinces, cantons, departments, regions and so on. Variations in population can be considerably large; e.g., in the United States, the population of California is $\sim$ 40 million whereas for Wyoming it is $\sim 570'000$. Parliamentary elections realized at the state level include elections to the United States Senate, as well as National and State Council elections in Switzerland. The United States presidential elections are a special case of elections whose outcome is determined by aggregating elections at the state level (except in Maine and Nebraska).

District: States can be further subdivided into districts, in such a way that these entities have more or less the same population. In parliamentary elections, each electoral district then elects its own, single representative. Examples thereof include the UK general elections, the United States House of Representatives elections.

Depending on the considered electoral process, results are gathered at one or the other level, and, if applicable, results in different units in that level are combined to give the outcome, i.e. the winning party.

2.2.2 Electoral Systems

There are multiple ways to select one or more representatives in an electoral unit based on the votes secured by the parties. In this chapter, we consider three main ones:

Proportional representation (PR): A set of m seats within an electoral unit are distributed among the parties to match their proportion of votes as closely as possible. The number of seats (m_A) attributed to party A with proportion of votes $\rho_A \in [0, 1]$ is then

$$m_A = \text{round}(\rho_A m)\,, \tag{2.9}$$

where the 'round' function rounds to the closest integer. For example, Switzerland elects the members of the one of the chambers of its Parliament through PR at the level called 'cantons'.

Single representative (SR): The electoral unit elects the candidate with majority of votes as its unique representative. It is a particular case of PR with $m = 1$. The United States House of Representatives are elected through SR at the level of districts. The second round of French presidential election follows the SR at the level of country.

Winner-Takes-All (WTA): All seats associated with an electoral unit are attributed to the party with the most votes. If there is only a single seat in the electoral unit, WTA reduces to SR. In the United States presidential election, the representatives are selected following the WTA process at the level of states.

The winning party of an election is the one that has the largest number of representatives (seats) at the level of the country.

2.3 Validation of the Model

Opinion dynamics models include many sociological assumptions and simplifications. These models could be used to analyze or evaluate a problem effectively only when they are validated against some historical data or experimental observation. The validation thus strengthens the results obtained while evaluating the problem.

In general, comparisons are indirect and statistical in nature, and an important part of the difficulty is translating a mathematical object, such as an agent's opinions, into an observed sociological quantity. It is important to emphasize that our model cannot predict or quantitatively model democratic electoral processes. Instead, it identifies and analyses electoral trends attributed to general social characteristics of voter populations or specific electoral systems.

Model validation requires a comparison between numerically obtained data and real-life measurement data. The scarcity of the latter that is necessary to compare with the numerically obtained data of interest from the models makes it harder to validate opinion dynamics models. The data of elections in the United States are particularly suited to compare the numerically obtained data from the model to historical data. It has primarily two major parties (Democrats and Republicans) and historical data is easily accessible.

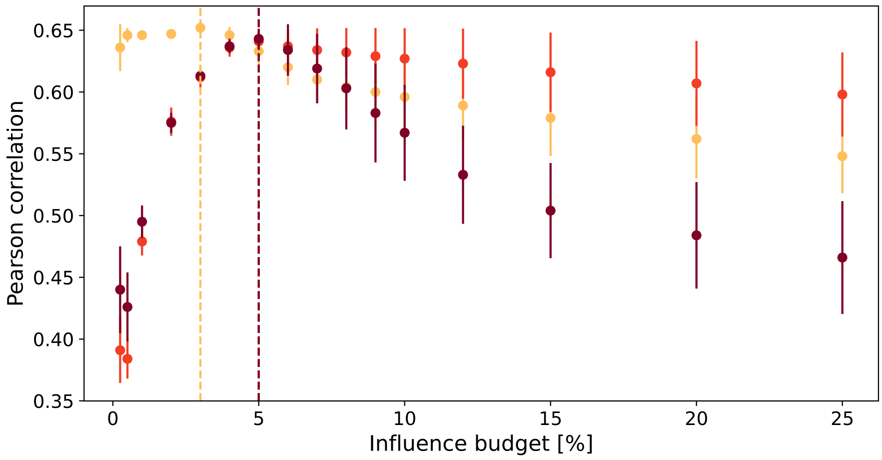


Figure 2.1: Pearson correlation coefficient between historical volatility and simulated volatility for three different distributions of natural opinions (yellow (unshifted polarized distribution with biased weight), orange (shifted polarized distribution with equal weights), dark red (shifted, unpolarized distribution)). Average is taken over different values of ϵ and error bars denote the standard deviation. The vertical dashed lines indicate the maximal correlation for each distribution of natural opinions.

For the validation of our model, we compare historical and numerical volatility of the House of Representatives elections in the United States, between 2012 and 2020, i.e., a complete sequence of elections in between two consecutive census-based redistrictings. Obviously, it is beyond the reach of any mathematical model to reproduce precise vote percentages of the Democrats or the Republicans within each district. However, it is possible to capture trends by observing how often the winning party changes from one election to the next in a given district.

We measure the historical volatility of the House of Representatives election in each district by counting the number of times the winning party (Democrats or Republicans) has changed from 2012 to 2020. A single representative is elected at the level of the districts for the House of Representatives elections. The more this number changes over time, the more volatile the electoral unit is. This historical volatility measure is illustrated in Fig. 2.2 **a**. Over the period 2012 - 2020, 78 districts changed the winning party at least once (See Fig. 2.2 **a**).

To compute the numerical volatility given by our model, we allocate an *influence budget* (IB) to influence a certain percentage of agents in the country, which is distributed evenly among the electoral units. The influence budget that we consider amounts to influencing $0.25\% - 20\%$ of the agents in the country. For a given value of $\epsilon \in [0.05, 0.8]$, we model each district with 501 agents and influence a given percentage of the population in each district in an attempt to change the winning party. We then simulate the election in each district and compute the

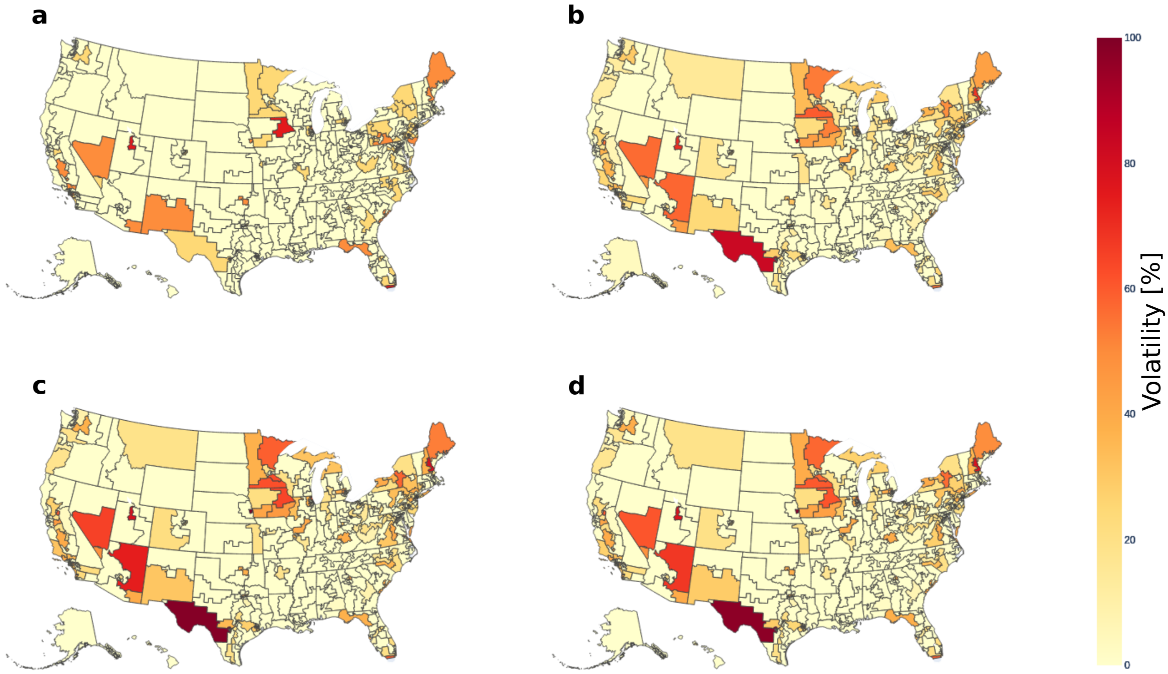


Figure 2.2: **a** Historical volatility of United States electoral district. The color scale on the map indicates the percentage of elections where the winning party changed from Republican to Democrats or vice versa, for the period between 2012 and 2020 for House of Representative elections. (Panels **b**, **c**, **d**) Numerical volatility of United States electoral district, computed as described in the text. The distribution of natural opinions corresponds to, unshifted polarized distribution with biased weight (**b**), shifted polarized distribution with equal weights (**c**), shifted, unpolarized distribution (**d**). There is a Pearson correlation coefficient of 0.65 between the historical and numerical volatility for all three different types of natural opinion distribution.

number of times this influence can change the outcome in each of these electoral units over 100 independent realizations of natural opinions. After influence, each district may or may not change the winning party. We repeat this process with the natural opinions generated from the historical data of the House of Representatives election from 2012 to 2020. For each value of confidence bound ϵ, averaging over all realizations of the number of times the influence has changed the winning party and election years gives our numerical volatility.

We consider three distributions of natural opinions (see Eq. (2.6)), where parameters are tuned to match the vote percentages given by the historical data:

Unshifted polarized distribution with biased weight : ($\mu = 0$, $\Delta = 0.5$, $\sigma = 0.2$), where the majority (given by the historic data) is obtained by tuning the size of each weight $p \in [0, 1]$;

Shifted polarized distribution with equal weights : ($\Delta = 0.5$, $\sigma = 0.2$) with balanced weights ($p = 0.5$), where the majority (given by the historical data) is obtained by biasing the whole population with $\mu \in \mathbb{R}$;

Shifted, unpolarized distribution : ($\Delta = 0$, $\sigma = 0.2, p = 0.5$) where the majority (given by the

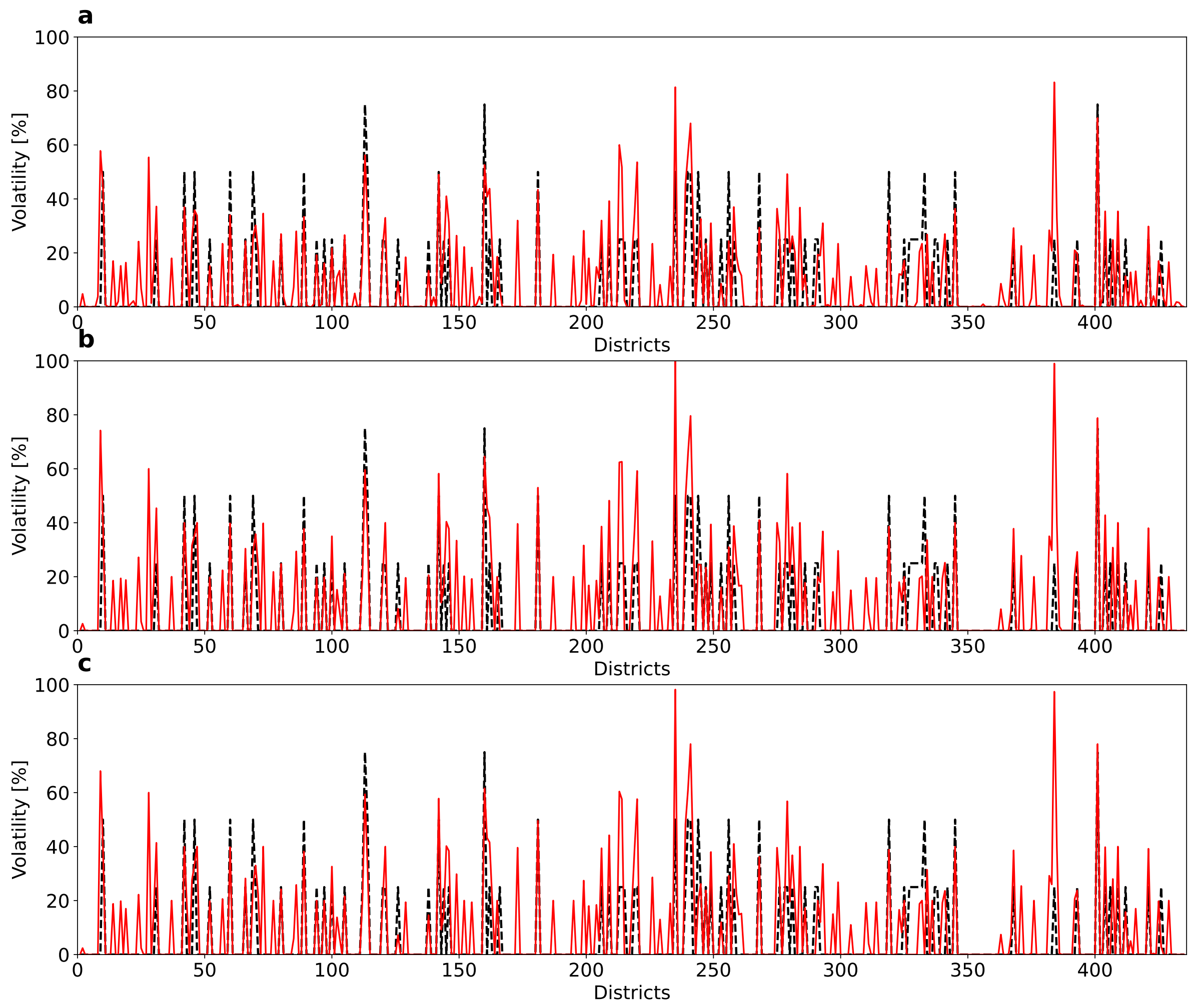


Figure 2.3: Comparison between historical volatility (black dashed line) and numerical volatility (red solid line) for 435 United States electoral districts numbered 1 to 435 (in alphabetical order on the horizontal axis) during House of Representative election from 2012 to 2020. The distribution of natural opinions corresponds to **a**: unshifted polarized distribution with biased weight, **b**: shifted polarized distribution with equal weights, **c**: shifted, unpolarized distribution.

historical data) is obtained by biasing the whole population through $\mu \in \mathbb{R}$.

We compare the results of the numerical volatility against historical volatility by computing the Pearson correlation coefficient between these quantities. The Pearson correlation coefficient averaged over confidence bound ϵ vs influence budget is shown in Fig. 2.1. The error bars in Fig. 2.1 denote the standard deviation of Pearson coefficient over confidence bound ϵ. The Pearson coefficient increases to a maximum value and then decreases as a function of influence budget. With higher values of influence budget we convince many agents which is unrealistic in actual elections, which leads to the aforementioned non-monotonous pattern. From this data, we determine the best Pearson correlation coefficient of around 0.66 and the corresponding confidence bound ϵ, for all three different types of natural opinion. This large Pearson coefficient indicates high cross-correlation between the historic data and our simulations. The influence budget at which Pearson coefficient reaches maximum value is 3% for the an unshifted polarized distribution with biased weight and 5% for the shifted polarized distribution with equal weights and shifted, unpolarized distribution of natural opinions. The effect of the confidence bound ϵ is minimal when the Pearson coefficient is at its maximum value, and its influence becomes more significant as the influence budget becomes larger. We conclude that our model captures the main trends and features of real elections.

The numerical volatility corresponding to the three distributions of natural opinions with the highest Pearson correlation coefficient is shown in Figs. 2.2 [**b-d**]. Note that for each type of distribution, we tune the confidence bound ϵ to maximize the Pearson correlation coefficient. Qualitatively, most of the districts that haven't changed their winning party over the period of 2012 to 2012 is also shown to be resilient to influence as per our model. Similarly results hold for the volatile electoral districts as well. Obvious differences exist, however, the two sets of data look overall similar, which is confirmed by a rather large Pearson correlation coefficient of $r \approx 0.66$ for all three different types of natural opinion. The large Pearson coefficient indicates that our model provides an accurate estimate of the robustness of electoral systems.

The Pearson correlation coefficient assesses the extent to which our numerical volatility aligns with historical data, gauging the strength and direction of similarity in the patterns of change between these datasets. However, it tends to overlook differences in the amplitude or magnitude of these changes. As a result, it may not adequately capture situations where the scale of variation between the two datasets differs significantly. To verify this, we plot Figs. 2.2 [**b-d**] differently by comparing the amplitude of numerical volatility with the historical volatility for each of the 435 United States electoral districts numbered 1 to 435 in the alphabetical order (see Fig. 2.3). Even the amplitude of numerical volatility shows an excellent fit to the historical volatility for all three different distributions of natural opinion, with minor differences for a few electoral districts. In the case of unshifted polarized distribution with biased weight, the numerical volatility values of electoral districts are slightly smaller than the other distributions, shifted polarized distribution with equal weights and shifted, unpolarized distribution. This behavior is a consequence of selecting an influence budget that leads to a higher Pearson correlation for the unshifted polarized distribution with biased weight distribution, specifically an IB of 3%. This influence budget choice is relatively smaller compared to the influence budget choices for the shifted polarized distribution with equal weights and shifted, unpolarized distributions, both of which are set at 5%. Given the low number of independent parameters in our model, we take this as a successful validation test for our model.

2.4 Robustness Analysis

In our model, the electoral process is modeled in two steps. In the first step, we model the opinion formation of agents based on the social characteristics of the population, such as bias in the population, a priori polarization of agents opinions and the openness of the agents towards the opinion of other agents (modeled by confidence bound ϵ). In the second step, we model different types of democratic electoral system that exist worldwide to elect the president of the country or for electing representatives for different branches of government such as executive, legislative and judiciary.

In order to understand the effect of each sociological aspects of population on the electoral systems to external influences, we identify each aspect independently on a single electoral unit.

First we consider the aspect of no bias ($p = 0.5, \mu = 0$) on the voter population, and later we introduce some bias keeping all other aspects fixed. The effect of sociological aspects on different electoral systems may differ, and we determine which electoral system is more robust and least robust to external influences. As we validated that our model gives an accurate estimate of the robustness of an election outcome, we investigate these aspects in this section based on our model.

2.4.1 Robustness Against Polarization and Bias

We first focus on a single electoral unit using the SR electoral system. The distribution of natural opinions within this electoral unit given by Eq. (2.6) is considered close to parity, meaning that a moderate effort would be sufficient to change the election outcome. With such a consideration, we can independently consider each of the sociologically aspects of the population. Following the strategy outlined in Section 2.1.2, we determine the minimal effort required to change the election outcome.

For our simulations, we consider 2001 agents in a single electoral unit and vary the confidence bound in the range of $\epsilon \in [0, 1.5]$. The odd number of agents in the electoral unit ensures that we obtain a majority of at least one agent for one of the parties. We study the effect of different parameters of the natural opinion distribution by computing the minimal effort across a range of (i) polarization parameter (Δ), (ii) global shift (μ), (iii) weight of Gaussian peak (p) of Eq. (2.6) as a function of confidence bound (ϵ).

In Fig. 2.4, we show three plots, each corresponding to a different interval of the polarization parameter. Specifically, we explore polarization parameter in the range $\Delta \in [0, 1]$ as a function of confidence bound $\epsilon \in [0, 1.5]$. It is important to note that the distribution of natural opinions becomes highly polarized when the polarization parameter $\Delta > 0.6$ and, our primary focus is on democratic elections characterized by low to moderate polarization. Within the range $\Delta \in [0, 0.6]$, we refer to the polarization parameter as the *democratic polarization parameter*. Fig. 2.5 shows an example of the final opinions of the agents at different values of confidence bound ϵ. We have chosen one polarization parameter value Δ from each interval of Fig. 2.4 to interpret our results within the corresponding interval.

For weak polarization parameter Δ, it is evident from Fig. 2.4 **a** that large values of opinion polarization within the range $\Delta \in [0, 0.6]$ is associated with a less robust election outcome and the effort required to alter the election outcome increases as a function of confidence bound ϵ. With increase in polarization, the density of agents close to neutral opinion ($x = 0$) decreases and this reduces the connectivity between the agents near the neutral opinion. This reduces the rigidity of change in opinion of agents, potentially leading to a change in their voting behaviour and the effort needed to change the election outcome decreases. Similarly with increase in confidence bound ϵ, the connectivity between the agents increases and the rigidity of change in opinion increases and a significant effort is necessary to manipulate the election result. Thus the effort needed to change the election outcome increases as a function of confidence bound ϵ.

At larger values of polarization parameter $\Delta \in [0.71, 0.85]$ as shown in Fig. 2.4 **b**, we observe a deviation in the behaviour of the effort needed to change the election outcome as a function of the confidence bound ϵ. Initially, as the confidence bound increases, we observe an increase in the effort required to change the election outcome. With moderate values of $\epsilon \in [0.1, 0.4]$, agents tend to move towards other agents with similar opinions, thus the density of agents near the neutral opinion ($x = 0$) are initially depleted. This leads to forming two distinct groups with similar opinions, making it more difficult to influence agents across group boundaries. The

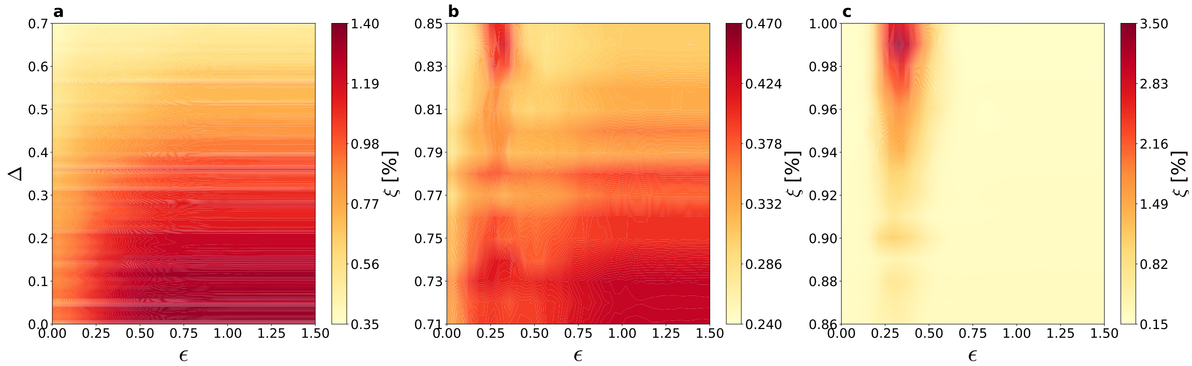


Figure 2.4: Average effort needed to change the outcome of an election with respect to polarization parameter Δ of the opinion distribution given by Eq. 2.6 and the confidence bound ϵ. The average is taken over 500 realizations of natural opinion with 2001 agents in single electoral unit. With $\sigma = 0.2$, $\mu = 0$, and $p = 0.5$, the interval of polarization parameter Δ is varied in the panels as follows, **a**: $\Delta \in [0, 0.7]$, **b**: $\Delta \in [0.71, 0.85]$, **c**: $\Delta \in [0.86, 1]$

formation of these two distinct groups is seen Fig. 2.5 **b**, as the gap between agents belonging to the two parties near the neutral opinion.

However, for larger values of $\epsilon > 0.4$, agents increase their interactions, and combined with attractive dynamics, and the difference between their opinion decreases. In this case, influencing agents becomes easier due to the reduced separation between their opinions, thus the effect is reversed with longer-range of interaction. Interestingly, even within this polarization parameter interval of $\Delta \in [0.71, 0.85]$, we observe that higher polarization of opinions is associated with a less robust outcome. This means that as the opinions become more polarized, the election outcome becomes more susceptible to change. This interval serves as a transition range for the effort needed to change the election outcome curve as it shifts from a monotonous trend to a non-monotonous one, which is observed in the next polarization parameter range (see Fig.2.4 **c**).

Fig. 2.4 **c** represents the effort needed to change the election outcome for still higher polarization parameter $\Delta \in [0.86, 1]$ as a function of confidence bound ϵ. We observe a different behaviour for the effort needed to change the election outcome compared to the previous cases. As the confidence bound ϵ increases, we observe a strong non-monotonous pattern in the effort required to change the election outcome, and the relationship between the effort and the confidence bound becomes more complex.

For smaller values of confidence bound parameter $\epsilon \in [0, 0.2]$, the agents tend to follow their natural opinion as the interactions are minimal. With only a few agents near the neutral opinion and minimal interactions, it is easier to change the election outcome (see Fig. 2.5 **c**).

In the interval $\epsilon \in [0.2, 0.6]$, with higher values of the polarization parameter, the system tends to form two well-separated groups of agents, each corresponding to one of the parties (see Fig. 2.5 **c**). The distance between these two groups increases with the polarization parameter, and influencing agents to switch their preferences to the other party increasingly difficult. Consequently, the effort needed to change the election outcome increases as the separation between the two groups becomes more pronounced. Agents in one group are less likely to be influenced by agents in the other group due to the substantial differences in opinions. In this

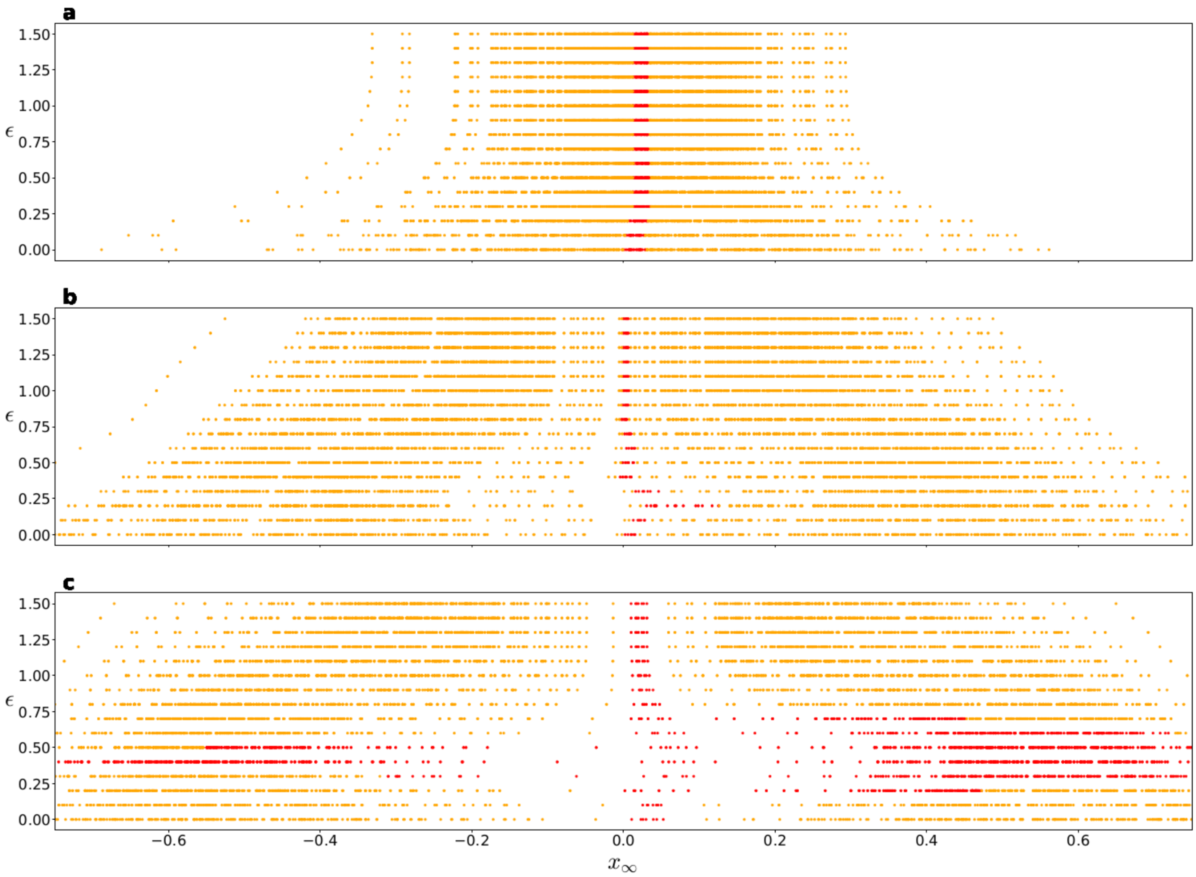


Figure 2.5: Final opinions of the agents (x_∞) with different values of confidence bound parameter ϵ for individual solutions given by Eq. (2.5). The final opinions of agents in the system are denoted by orange and red dots on the x-axis, and the corresponding y-axis denotes the confidence bound parameter ϵ used to construct the interaction network. The agents colored in red indicate the agents influenced to change the election outcome. The polarization parameter Δ is varied in the panels as follows, **a**: $\Delta = 0$, **b**: $\Delta = 0.75$, and **c**: $\Delta = 1$.

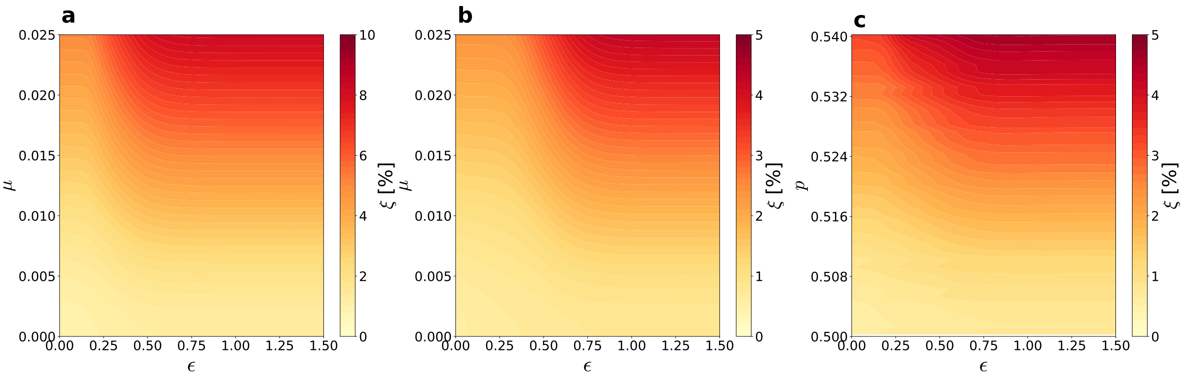


Figure 2.6: Average effort needed to change the outcome of an election with respect to bias (μ and p) of the natural opinion distribution given by Eq. (2.6) and the confidence bound ϵ. The average is taken over 500 realizations of natural opinions with 2001 agents in single electoral unit. **a**: Average effort as a function of the confidence bound ϵ and bias μ, for a Gaussian distribution ($\Delta = 0.0, p = 0.5$). **b**: Average effort as a function of the confidence bound ϵ and bias μ, for a bigaussian distribution ($\Delta = 0.5, p = 0.5$). **c**: Average effort as a function of the confidence bound ϵ and the balance parameter p, for a bigaussian distribution ($\Delta = 0.5, \mu = 0.0$).

interval of ϵ, a higher polarization of the opinion is related to more robust outcome.

However, as the confidence bound ϵ continues to increase, the agents interact more frequently. This increased interaction reduces the distance between two groups and tries to form a single large group of agents near the average opinion of the population, i.e. neutral opinion (see Fig. 2.5 **c**). In this case, it becomes easier to influence agents to change their vote preferences as the agents are more interconnected and the distance between groups diminishes. As a result, the effort needed to change the election outcome decreases for larger values of the confidence bound, resulting in a non-monotonic behaviour for the effort curve as a function of ϵ in the case of higher polarization parameter values.

In Fig. 2.6, we examine the effort required to alter the election outcome based on different parameters (μ, Δ, p) of the opinion distribution given by Eq. (2.6), as a function of confidence bound ϵ. From Figs. 2.6 **a** and 2.6 **b**, it is evident that shifting the mean of the opinion distribution towards one side increases the effort needed to change the election outcome. This observation aligns with our expectations since greater effort is required when one party holds a significant majority, regardless of the level of polarization. In other words, when there is a large imbalance in the distribution, the number of agents to be influenced to change the election outcome will be increase, and influencing a substantial number of agents becomes more challenging. As anticipated, Fig. 2.6 **c** demonstrates that the robustness of the election outcome, as measured by the effort needed to change it, increases with both the confidence bound and the discrepancy between the sizes of the two parties. When one party has a larger group of voters, more agents must be influenced to swing the result. Additionally, agents who are more connected within the interaction network are generally harder to persuade, hence the higher robustness associated with larger confidence bound ϵ. This indicates that increasing the confidence bound raises the difficulty in changing the opinions of well-connected agents, contributing to a more resilient election outcome. We concentrate only on the democratic polarization parameter range for our study on robustness of different electoral systems.

2.4.2 Robustness of Electoral Systems

Electoral systems are designed to elect representatives based on voters' preferences. The design of an electoral system can significantly impact the outcome of an election and the representation of different groups within society. Evaluating the robustness of different electoral systems helps ensure the system is fair, transparent, and representative. We evaluate the robustness of different electoral systems taking into account the population's diversity of the country by tuning the parameters of natural opinion given in Eq. (2.6).

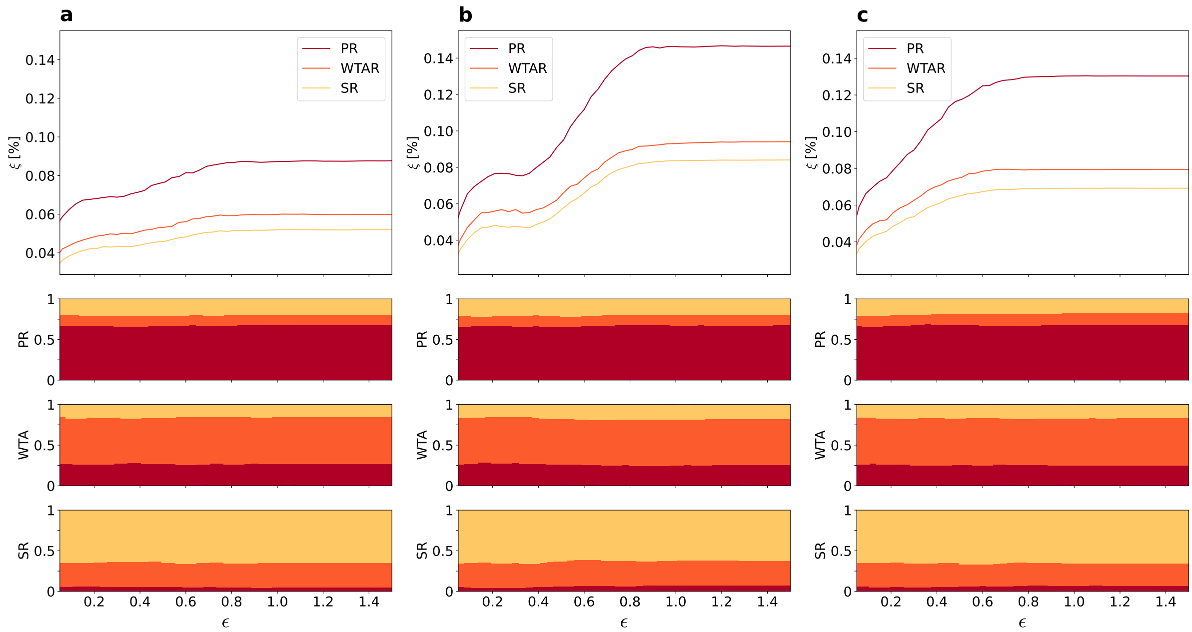


Figure 2.7: Robustness of different electoral systems with respect to the confidence bound, for three different distributions of natural opinions. Simulations are realized over random synthetic countries with multiple electoral units. For each realization, a random difference of votes (uniform between 0% and 10%) is introduced following **a**: unshifted polarized distribution with biased weight, **b**: shifted polarized distribution with equal weights and **c**: shifted, unpolarized distributions of natural opinions. In each panel, the top figure shows the effort (averaged over 1500 realizations of natural opinion) necessary to change the outcome of an election, as a function of the confidence bound ϵ. Each curve corresponds to one of the three electoral systems: single representative (yellow), winner-takes-all (orange), and proportional representative (dark red). The bottom plots show the proportion of times each electoral system was the most robust (red), second robust (orange), and least robust (yellow).

To compare the robustness of different electoral systems, we need to fix an electoral structure (country, states, districts) and the distribution of opinions in the population. We first compute the election outcome following different electoral systems and then introduce an external influence in attempt to change it.

For our simulations, we construct a set of 15 synthetic countries, each consisting of $n_s \in [16, 20]$ states. We explore three different seat distributions for each country with n_s states. The number of seats per state is drawn from the interval $[3, 15]$, with an average of 9 seats per state. The assignment of seats within each state is done arbitrarily. The number of agents in each state is proportional to the number of seats in the corresponding state.

To make the election more realistic, a partisan bias is introduced, which randomly favors one or the other party in each state. The percentage of initial bias of each party is randomly generated, and the maximum difference of votes between the parties is 10%. This allows us to capture a range of scenarios in which parties may have differing levels of support. Using these parameters, we assess the robustness of three different electoral systems (PR, SR, and WTA), as discussed in Section 2.2. By conducting numerical simulations, we can compare the robustness of different electoral systems across multiple hypothetical countries, providing insights into their performance under diverse conditions.

For our simulations, we consider 101 agents in each district, and one seat is allotted for each district representation within the corresponding state. Thus the total number of seats in each state matches the number of districts within the corresponding state. This ensures a consistent and comparable representation of the population within each state. For our simulation, the interaction network is constructed at the state level. This means that agents within a state interact with each other, and there is no interaction among the agents in different states. We conduct the elections at the level of states and the outcomes of elections are aggregated at the country level for all the three different types of electoral systems.

In each state, the percentage of agents supporting each party as per the distributions of natural opinions is calibrated to align with the bias introduced in the corresponding state. We perform multiple realizations for each combination of the electoral system, natural opinion distributions, and synthetic country. Since it is not possible to have individual realizations of elections, the multiple realizations of synthetic elections are used to extract general trends of actual elections. By averaging the efforts required to change the election outcome across the 15 different synthetic countries and the realizations of natural opinion, we obtain the robustness of each electoral system. This approach is robust and statistically relevant as it considers multiple factors such as seat distributions, opinion distributions, and synthetic countries.

In Fig. 2.7, we present the results of our simulations, showcasing the average effort needed to change the election outcome as a function of the confidence bound ϵ. Additionally, we display the proportion of realizations where each electoral system was the most robust, second most robust, or least robust. Based on the findings illustrated in Fig. 2.7, we observe a consistent pattern regardless of the distribution of natural opinions, including the nature of the random shift giving the initial electoral bias, or the confidence bound between agents. The PR system of elections are consistently the most robust, demonstrating a higher resistance to external influences. However the SR system is the least robust to external influences. These results highlight the advantages of PR regarding resilience to manipulation and the ability to reflect the diversity of opinions within a population. The WTA system, although less robust than PR, still offers moderate resistance to external influences. On the other hand, the SR proves to be the least robust, suggesting its susceptibility to manipulation and external influences.

From Fig. 2.7, we gain further insights into their relative performance by considering the proportion of realizations where each electoral system was the most, second most, or least robust. Beyond the average robustness, the most resilient electoral system is the PR system in about 65% of the 1500 statistical realizations of synthetic countries we considered, the WTA system in about 25% and the SR in less than 10% of those cases. This reinforce the conclusion that PR is the most robust electoral system among the evaluated ones.

Indeed, our observations align with the findings in Fig. 2.4 and 2.6. As illustrated in those figures, increasing the confidence bound ϵ increases the system's robustness within the democratic polarization parameter. A population becomes more resilient to external influence and manipulation by promoting communication and interaction among agents. This is reflected in the increased robustness observed as the confidence bound ϵ increases. When agents have more opportunities to exchange and share their opinions, it becomes more challenging for external forces to sway the overall outcome of the election.

Additionally, among the three types of opinion distributions, we find that shifted polarized distribution with equal weights exhibits slightly higher robustness to external influence than the other two distributions (unshifted polarized distribution with biased weight, shifted, unpolarized distribution). Overall, these observations further emphasize the importance of communication and promoting democratic participation in maintaining the robustness of an electoral system.

2.5 A case study on United States election data

We verify that the results obtained for the robustness of electoral system holds for some historical data. We use the electoral structure of the United States, on which we compare different electoral systems at the level of states. The number of seats in each state in the United States is proportional to the population of each state, and we use the same number of seats in the United States for our simulations.

Natural opinions for agents in the electoral units are generated using the historical election results of House of Representative elections from the year 2012 to 2020. We use the results of each of these elections as bias or shift in the opinion distribution. For simulations at the level of states, we consider 101 agents per district, which yields states with populations ranging from 101 agents in the state of Wyoming to 5353 agents in the state of California. The interaction networks for our simulations are constructed at the level of states.

We compare the electoral systems over the three distributions of natural opinions discussed in Sec. 2.3: unshifted polarized distribution with biased weight, shifted polarized distribution with equal weights, and shifted, unpolarized distribution. For generating the natural opinion of agents corresponding to each type of distributions, we proceed at the level of districts. For each district we tune the distributions of natural opinions such that they agree with the historical data of HOR electoral year from 2012 to 2020 (5 elections). The natural opinions were gathered among the districts composing the state, and the agents interact within the state level. For each HOR electoral year, we generate 100 individual realizations of natural opinion corresponding to each of the three distributions of natural opinions. Finally, we average the effort over all years and realizations for all the three opinion distributions for each of the different electoral systems.

We show in Fig. 2.8, irrespective of the opinion distribution, we observe same pattern as in Fig. 2.7, where the PR elections are by far the most robust, followed by WTA elections and the SR being the least robust system. We also observe that increasing the confidence bound ϵ increases the system's robustness. The simulations shows little difference with the synthetic countries, in terms of the magnitude of effort needed to change the election's outcome and the proportion of times the WTA was most robust. The magnitude of effort needed to change the election outcome is indeed different because of the electoral bias corresponding to Republicans and Democrats in the elections. The WTA system that appears to never be the most robust in this case.

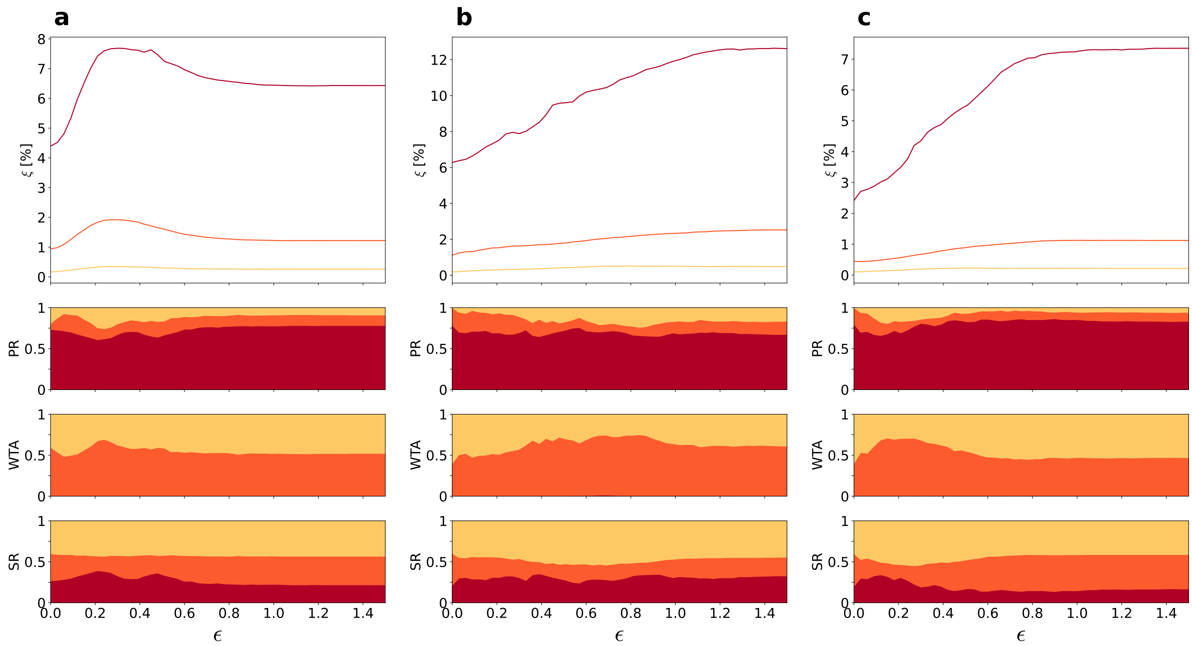


Figure 2.8: Robustness evaluation on a simulation of the United States House of Representative elections, between 2012 and 2020 (100 realization for each even year). Elections are performed at the level of states, with results aggregated from the districts. Top row: average effort needed to change the election outcome for three electoral systems (PR: red, WTA: orange, and SR: yellow) and three distributions of natural opinions (unshifted polarized distribution with biased weight, shifted polarized distribution with equal weights, and shifted, unpolarized distribution from left to right). Historical election results were introduced in the natural opinion distribution via parameters p for unshifted polarized distribution with biased weight, and μ for shifted polarized distribution with equal weights and shifted, unpolarized distribution. Bottom rows: Proportion of times each electoral system was the most robust (red), second robust (orange), and least robust (yellow).

However, the trend of electoral system remains the same regardless of the electoral structure and difference in vote percentage. Thus, the results obtained for a synthetic case holds for real elections which verifies our results to some historical data.

2.6 Conclusion

We concentrated on two-party system of elections. We have constructed an opinion dynamics model given by Eq. (2.4) to evaluate the robustness of different electoral systems to external influences. Our model is not meant to predict, nor to quantitatively model democratic electoral processes, but rather to extract electoral trends that can be attributed to general social characteristics of voter populations or to specific electoral modalities. In that spirit, the model is qualitatively validated as it reproduces the observed volatility of the US House of Representatives elections from 2012 to 2020, with a high Pearson correlation coefficient of 0.65

between them. This is quite remarkable, in particular given the notorious difficulty to validate results in computational social science.

Our results shows that a system is systematically more robust as the confidence bound ϵ, which drives agent connectivity, increases. These results emphasize in particular the need to encourage public debates during political campaigns, to strengthen democratic processes. We also observe that the robustness of an election outcome is favored by less polarization and biasing of the population. Surprisingly, in the democratic polarization parameter range, a polarized society is less robust to external influence than a less polarized society.

Our study compared the robustness of three different electoral system and we observed that proportional representation systems are more robust than winner takes all systems, which are more robust than those with a single representative. Among the three different types of distribution we have considered, shifted polarized distribution with equal weights seems to be more robust and unshifted polarized distribution with biased weight seems to be less robust to external influence compared to corresponding other two distributions.

There are, of course, many directions to extend the work. One of the most appealing is extending the model to more than two parties. Such an extension would allow modelling similarities between parties, creating a coalition, and abstention in voting, which is explained in the next chapter.

2.7 Appendix

Lemma 4. *The matrix* $M := D_\epsilon^{-1} L_\epsilon + \mathbb{I}$ *(used in Sec. 2.1) has a real positive spectrum.*

Proof. First, let us compute the eigenvalues of $D_\epsilon^{-1} L_\epsilon$, which are the zeros of the characteristic polynomial

$$\begin{aligned} 0 &= \det\left(\lambda\mathbb{I} - D_\epsilon^{-1} L_\epsilon\right) \\ &= \det\left[D_\epsilon^{-1/2}\left(\lambda\mathbb{I} - D_\epsilon^{-1/2} L_\epsilon D_\epsilon^{-1/2}\right) D_\epsilon^{1/2}\right] \\ &= \det\left(\lambda\mathbb{I} - D_\epsilon^{-1/2} L_\epsilon D_\epsilon^{-1/2}\right) . \end{aligned} \tag{2.10}$$

We then see that the spectrum of $D_\epsilon^{-1} L_\epsilon$ coincides with the spectrum of $D_\epsilon^{-1/2} L_\epsilon D_\epsilon^{-1/2}$. The matrix $D_\epsilon^{-1/2} L_\epsilon D_\epsilon^{-1/2}$ being real symmetric, its spectrum is real. Furthermore, if λ is an eigenvalue of $D_\epsilon^{-1} L_\epsilon$ then $\lambda + 1$ is an eigenvalue of M. The spectrum of M is then real.

Second, since $M_{ii} = 2$ and $\sum_{j \neq i} |M_{ij}| = 1$ for all i, By Gershgorin's Circles Theorem [160], the eigenvalues of M have real part greater than 1, which concludes the proof. □

Lemma 5. *Ordering agents with respect to their natural opinion,* x_i^0*, or by their final opinion,* $(x_\infty)_i$ *gives the same ordering.*

Proof. Assume that at time $t = 0$, all agents have opinion equal to their natural opinion, $\boldsymbol{x}(0) = \boldsymbol{x}^0$. Then it suffices to prove that the dynamics generate no crossing between agents. We will proceed by contradiction.

Assume that there are crossings for the agent's opinion and that the first one occurs at time t, between agent 1 and 2, i.e. $x_1(t) = x_2(t)$ and all other agents are ordered according to their *natural opinions*. Without loss of generality, we assume $x_1(0) < x_2(0)$. Before going further, let us define N_i, the set of neighbors of agent i and

$$\begin{aligned} N_{ij} &= N_i \cap N_j, \\ d_{ij} &= |N_{ij}|, \\ N_{i\setminus j} &= N_i \setminus N_j, \\ d_{i\setminus j} &= |N_{i\setminus j}| = d_i - d_{ij}. \end{aligned} \tag{2.11}$$

By the definition of interaction network,

$$\max_{i \in N_{1\setminus 2}} x_i(0) \leq \min_{i \in N_{12}} x_i(0) \leq \max_{i \in N_{12}} x_i(0) \leq \min_{i \in N_{2\setminus 1}} x_i(0), \tag{2.12}$$

as illustrated in Fig. 2.9.

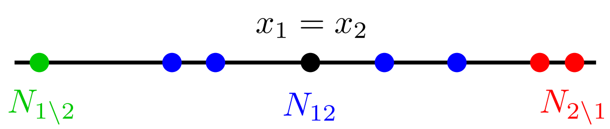


Figure 2.9: Illustration of the neighbors of 1 and 2 at time t

The time derivative of x_1 and x_2 at time of crossing

$$\begin{aligned}
\dot{x}_1 &= -\frac{1}{d_1}\sum_{j\in N_1}(x_1 - x_j) - (x_1 - {x_1}^0) \\
&= -\frac{d_{1\backslash 2}}{d_1}\underbrace{\Big[\frac{1}{d_{1\backslash 2}}\sum_{j\in N_{1\backslash 2}}(x_1 - x_j)\Big]}_{\Sigma_1} - \frac{d_{12}}{d_1}\underbrace{\Big[\frac{1}{d_{12}}\sum_{j\in N_{12}}(x_1 - x_j)\Big]}_{\Sigma_{12}} - (x_1 - {x_1}^0) \\
&= -\Big[\frac{d_1 - d_{12}}{d_1}\Sigma_1 + \frac{d_{12}}{d_1}\Sigma_{12}\Big] - (x_1 - {x_1}^0),
\end{aligned} \tag{2.13}$$

and similarly (recall that we assumed $x_1 = x_2$)

$$\dot{x}_2 = -\Big[\frac{d_2 - d_{12}}{d_2}\Sigma_2 + \frac{d_{12}}{d_2}\Sigma_{12}\Big] - (x_2 - {x_2}^0). \tag{2.14}$$

Now, as we assumed t to be the time of first opinion crossing, Eq. (2.12) is still satisfied. In particular $\sum_1 \geq \sum_{12} \geq \sum_2$ and

$$\frac{d_1 - d_{12}}{d_1}\Sigma_1 + \frac{d_{12}}{d_1}\Sigma_{12} \geq \Sigma_{12} \geq \frac{d_2 - d_{12}}{d_2}\Sigma_2 + \frac{d_{12}}{d_2}\Sigma_{12} \tag{2.15}$$

Also, by assumption,

$$x_1(t) - {x_1}^0 \geq x_2(t) - {x_2}^0 \tag{2.16}$$

Plugging Eqs. (2.15) and (2.16) into Eqs. (2.14) and (2.13) yields $\dot{x}_1 \leq \dot{x}_2$. Agents 1 and 2 will then avoid crossing, contradicting our assumption that t was the time of first crossing, which concludes the proof. □

Lemma 6. *The matrix $M^{-1} := (D_\epsilon^{-1}L_\epsilon + \mathbb{I})^{-1}$ is row-stochastic.*

Proof. We need to show the two following inequalities

$$0 \leq \left(M^{-1}\right)_{ij} \leq 1\,, \tag{2.17}$$

and that

$$\sum_j \left(M^{-1}\right)_{ij} = 1\,, \qquad\qquad \forall\, i\,. \tag{2.18}$$

Note first that one can rewrite

$$M^{-1} = \left(D_\epsilon^{-1}L_\epsilon + \mathbb{I}\right)^{-1} = \left(L_\epsilon + D_\epsilon\right)^{-1} D_\epsilon\,. \tag{2.19}$$

The matrix $L_\epsilon + D_\epsilon$ is real, symmetric, and by Gershgorin's Circle Theorem [161] it is positive definite. It has non-positive off-diagonal elements. By definition, it is then a Stieltjes matrix and its inverse is *non-negative* [162, 163] . Equivalently, for all i, j,

$$\left(M^{-1}\right)_{ij} = \left[\left(L_\epsilon + D_\epsilon\right)^{-1} D_\epsilon\right]_{ij} \geq 0\,, \tag{2.20}$$

which proves the first inequality in Eq. (2.17). Let $\mathbf{1}$ be the vector of all ones of dimension N, and Eq. (2.18) is proven by direct computation:

$$\begin{aligned}\mathbf{1} &= \left(D_\epsilon^{-1}L_\epsilon + \mathbb{I}\right)^{-1}\left(D_\epsilon^{-1}L_\epsilon + \mathbb{I}\right)\mathbf{1} \\ &= \left(D_\epsilon^{-1}L_\epsilon + \mathbb{I}\right)^{-1}\left(\underbrace{D_\epsilon^{-1}L_\epsilon\mathbf{1}}_{=0} + \mathbf{1}\right) \\ &= M^{-1}\mathbf{1}\,.\end{aligned} \tag{2.21}$$

Rewriting Eq. (2.21) component-wise, one gets Eq. (2.18), which, together with Eq. (2.20) proves the second inequality in Eq. (2.17) and conclude the proof. □

Proposition 7. *Let* $\boldsymbol{x}^0 \in [-1,1]^N$ *and* $\tilde{\boldsymbol{x}} = \boldsymbol{x}^0 + \boldsymbol{\mu}$ *where* $\boldsymbol{\mu} \in \mathbb{R}^N$ *. Then,*

$$\tilde{\boldsymbol{x}}_\infty = \boldsymbol{x}_\infty + \boldsymbol{\mu}, \tag{2.22}$$

where $\boldsymbol{x}_\infty = (D_\epsilon^{-1}L_\epsilon + \mathbb{I})^{-1}(\boldsymbol{x}^0 + \boldsymbol{\omega})$ *and* $\tilde{\boldsymbol{x}}_\infty = (D_\epsilon^{-1}L_\epsilon + \mathbb{I})^{-1}(\tilde{\boldsymbol{x}} + \boldsymbol{\omega})$.

Proof.

$$\begin{aligned}\tilde{\boldsymbol{x}}_\infty &= (D_\epsilon^{-1}L_\epsilon + \mathbb{I})^{-1}(\tilde{\boldsymbol{x}} + \boldsymbol{\omega}) \\ &= (D_\epsilon^{-1}L_\epsilon + \mathbb{I})^{-1}(\boldsymbol{x}_0 + \boldsymbol{\mu} + \boldsymbol{\omega}) \\ &= (D_\epsilon^{-1}L_\epsilon + \mathbb{I})^{-1}(\boldsymbol{x}_0 + \boldsymbol{\omega}) + (D_\epsilon^{-1}L_\epsilon + \mathbb{I})^{-1}(\boldsymbol{\mu}) \\ &= \boldsymbol{x}_\infty + \boldsymbol{\mu}, \quad \text{as } (D_\epsilon^{-1}L_\epsilon + \mathbb{I})^{-1} \text{ is row-stochastic as proved in Lemma 6.}\end{aligned} \tag{2.23}$$

□

2.7.1 Influence strategies on single electoral unit

We investigated (numerically) two strategies, namely, *random*, and *minimum*, described below.

- **Random strategy**: We chose the agents to influence uniformly at random. This strategy is not expected to be very efficient, but it gives a baseline to compare other strategies and allows us to get some statistics after multiple realizations.

- **Minimum strategy**: Here we target those agents who are undecided or neutral. One can check that, if $\mathbb{L}_\epsilon^u$ is constructed as in Eq. (2.2), the order of the final opinions x_∞^u on the real axis is the same as the order of natural opinions (See Lemma 5). One can then suppose that the easiest agents to influence are the ones that are the closest to opinion zero. This strategy then influences first the agents whose natural opinion is smallest positive value. This is as well intuitively reasonable as it is easier to change the opinion of such an agent due to their low opinion strength.

We confirm our intuition that Minimum strategy is more efficient (low-effort) in changing the election outcome as shown in Fig. 2.10. Targeting agents who are undecided or neutral can be more effective than trying to sway voters who already have a strong opinion.

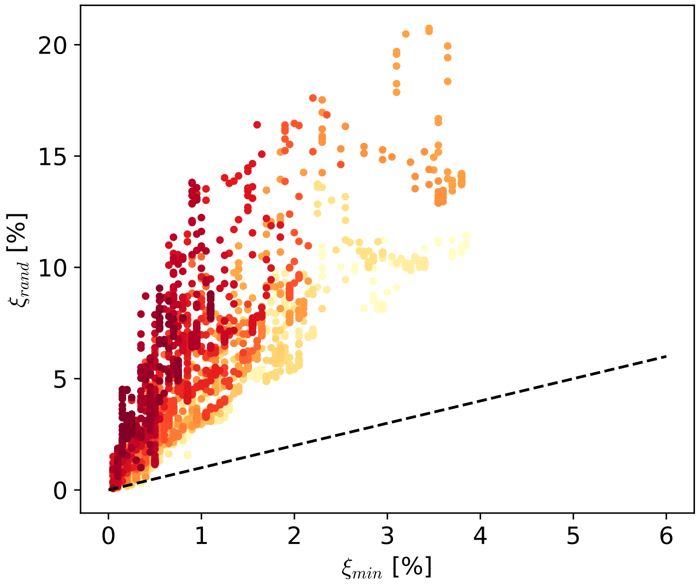


Figure 2.10: The effort needed to change the election outcome following *Random strategy* vs *Minimum strategy*. Each point denotes a natural opinion with different confidence bound ϵ, polarization Δ and bias μ. The black dotted line denotes the identity function for comparison.

2.7.2 Influence strategies for multiple electoral units

When a country is composed of multiple electoral units, there are different ways to prioritize the electoral units to be influenced first. We want to investigate the strategies that change the elections with as few agents targeted as possible. Here we investigate two strategies, namely *Minimum majority* and *Minimum population* as described below.

- **Minimum majority**: We target electoral units with the smallest majority. When the majority is small, even a change of opinion of a few agents in the population can have a significant impact on the final result.

- **Minimum population**: We target electoral units with the smallest population. Smaller electoral units have fewer agents and changing the outcome in such a unit requires to influence less agents.

Based on our calculations, it appears that targeting electoral units with the smallest difference in votes (minimum majority) is the effective strategy to change the outcome of an election (see Fig. 2.11). This strategy requires the external influence to shift only a relatively small number of votes to change the election outcome. As aforementioned in the main discussion, we have adopted minimum majority strategy which requires minimum effort to change the election outcome for evaluating the robustness of different electoral systems.

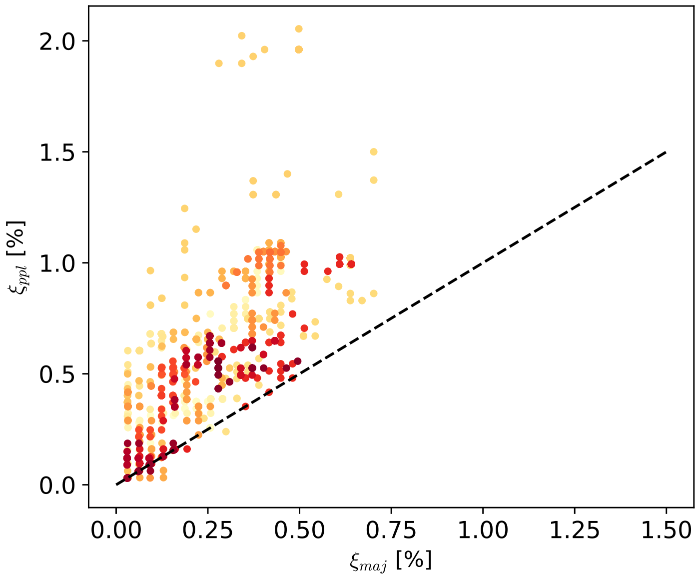


Figure 2.11: The effort needed to change the election outcome following *Minimum Population* vs *Minimum Majority* strategies with 7 states. We consider each state to have $400 - 500$ agents. Here each point denotes natural opinion with different values of confidence bound ϵ, polarization Δ and bias μ. The black dotted line denotes the identity function for comparison.

3 Multi-party System

Most of the democratic countries have more than two parties, for example countries such as Switzerland, Germany, India, etc. In this chapter, we extend our model to include the representation of multiple political parties in a legislative body or governing body. Unlike systems favoring two major parties, a multi-party system allows a broader range of political parties to participate and have their voices heard. This can lead to a more inclusive and representative democracy, where a wider spectrum of opinions and perspectives are considered in decision-making processes.

A primary challenge in expanding our model lies in accommodating the inherent political ordering within the societal structure, specifically the left-right political spectrum. An agent who supports leftist ideologies will typically hold a considerably less favorable view of right-wing perspectives. The approbation of an agent towards parties is represented using an opinion vector. We include the inherent left-right political spectrum of the society in the model by introducing some constraints on the natural opinions of the agents. This is detailed in section 3.1.2. We assume that each agent casts their vote in favor of the political party that aligns most closely with their convictions.

A multi-party electoral system has different ways of electing the legislative body's representatives, which have a major role in the results of elections. Here we study the effect of the left-right political spectrum in a multi-party system and evaluate the relative robustness of different electoral systems. We describe below the model of opinion formation and formalise the various electoral systems we investigate.

3.1 Opinion Dynamics

We consider a multi-party electoral system with p parties and the opinion X_i of an agent i is a vector of dimension p. In a democratic electoral system, all agents are treated equally, thus the sum of the elements of opinion vector for each agent is the same. Thus we represent the opinion of each agent as a normalized vector on the *simplex*, S_{p-1}. The detailed description of opinion space is given in Sec. 3.1.1.

To each agent $i \in \{1, \cdots, N\}$, we associate a natural opinion $X_i^0 \in S_{p-1}$, which is the opinion that agent i would hold if they were not interacting with any other agent. Then X_{iq}^0 is the approbation that agent i gives to the q^{th} party. We consider the linear consensus dynamics with natural opinion $X^0 \in (S_{p-1})^N$.

In similar spirit to the two-party system discussed in Sec 2.1, we define the interaction network based on the confidence bound $\epsilon > 0$, using the 1-norm distance between two agents. Then instead of Eq. (2.3), the dynamics is given by

$$\dot{X}_{iq} = d_i^{-1} \sum_j a_{ij}(X_{jq} - X_{iq}) + (X_{iq}^0 + W_{iq} - X_{iq}) \,, \tag{3.1}$$

with W_{iq} being the external influence applied to agent i on the q^{th} party (see Sec. 3.1.3). In vector form, for an electoral unit u with n agents, Eq. (3.1) yields the linear dynamics

$$\dot{X}^u = \Big[(X^0)^u + W^u\Big] - \Big[(D^u_\epsilon)^{-1} L^u_\epsilon + \mathbb{I}\Big] X^u \,. \tag{3.2}$$

As discussed in the previous chapter, Eq. (3.2) is a linear system and the matrix $(D^u_\epsilon)^{-1}L^u_\epsilon + \mathbb{I}$ has strictly positive eigenvalues. Therefore, it is globally stable and exponentially converges toward the equilibrium

$$X^u_\infty = \Big[(D^u_\epsilon)^{-1} L^u_\epsilon + \mathbb{I}\Big]^{-1} \Big((X^0)^u + W^u\Big) \,, \tag{3.3}$$

which then gives the final opinion of agents. For two-party system, the extended model exactly reduces to the one-dimensional model discussed in Chap. 2 (see Appendix for equivalence with two-party system).

3.1.1 The Opinion Space

In a democratic system with $p \geq 2$ parties, we represent the opinion of each agent as a point in the $(p-1)$-*simplex*

$$S_{p-1} = \left\{ \boldsymbol{x} \in \mathbb{R}^p \ : \ \sum_{k=1}^{p} x_k = 1 \,, x_k \geq 0 \right\} \,, \tag{3.4}$$

whose vertices are $\boldsymbol{v}_i$, $i \in \{1, \cdots, p\}$ with all components equal to zero, except the i^{th} which equals one. The yellow, orange, and maroon areas in Fig. 3.1 represent the domains of the 2-simplex where agents would vote for party 1, 2, and 3 respectively.

Without loss of generality, we can assume that the indexing of the parties (from 1 to p) matches the left-right political spectrum, i.e., party 1 is the left-most party and party p is the right-most. In order to guarantee coherence of each agent's opinions, we do not allow opinion vectors where two parties that are far away on the left-right spectrum have a large value while the parties in between have a low value. We need the opinion vector to satisfy some form of monotonicity. An opinion vector $\boldsymbol{x} \in S_{p-1}$ is acceptable only if its components increase until the largest one and then decrease. Mathematically, if $x_m = \max_i x_i$, we impose

$$x_i \leq x_j \,,\ \forall i \leq j \leq m \,, \qquad\qquad x_i \leq x_j \,,\ \forall m \leq j \leq i \,, \tag{3.5}$$

reducing the domain of allowed opinion. This ordering leads to no interactions or weak interactions between the agents of left and right-most parties. For instance, for $p = 3$ parties, Eq. (3.5) cuts out half of the yellow and dark red areas in Fig. 3.1, removing the shaded parts of the admissible opinion domain.

Furthermore, in order to isolate the impact of opinions on the election outcomes, we focus our analysis on societies where all the parties have a similar level of approbation in the population. After the ordering imposed by Eq. (3.5), one realizes that the left-most and right-most parties have the smallest volume in the opinion space. The ordering leads to some artifacts in the outcome of elections due to the increased density of agents in the extreme parties, which need to be removed. A detailed discussion of the artifact of ordering is given in the Appendix 3.6.2 for reference. Therefore, we need to reduce the volume V_i of the other parties $i \in \{2, \cdots, p-1\}$ in order to match the volumes of parties 1 and p.

We reduce the volume of the *moderates* (i.e., non-extremists) by sliding the summit of the simplex defining their area ($\boldsymbol{v}_i$) towards the barycenter $\boldsymbol{c}$ of the simplex, reaching the new

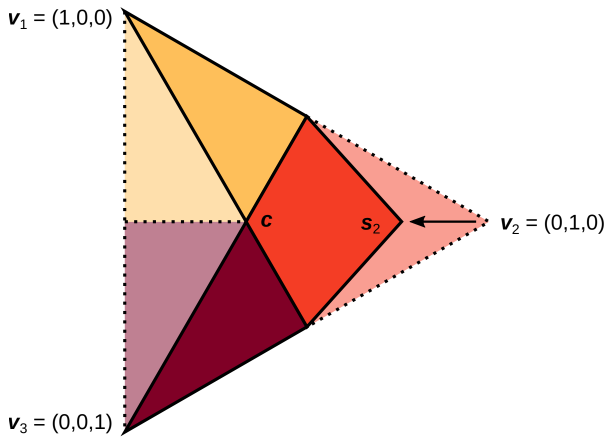


Figure 3.1: Admissible opinion domain in the 2-simplex (i.e., for three parties). The summits $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$ represent the pure opinions of each party respectively. The yellow and maroon shaded areas are removed from the admissible domain, because they represent inconsistent opinions, where $v_1 \geq v_3 \geq v_2$ or $v_3 \geq v_1 \geq v_2$. The shaded orange area is removed in order to balance the volume allocated to each party in the simplex, by sliding the summit $\boldsymbol{v}_2$ towards the barycenter $\boldsymbol{c}$ to $\boldsymbol{s}_2$.

summit $\boldsymbol{s}_i$ such that $V_i = V_1$ (see Fig. 3.1). In the Appendix 3.6.3, we show that the position of $\boldsymbol{s}_i$ can be determined in closed form and rather elegantly as

$$\boldsymbol{s}_i = \gamma_i \boldsymbol{c} + (1 - \gamma_i)\boldsymbol{v}_i \,, \qquad \gamma_i = \frac{1}{\binom{p-1}{i-1}} \,. \tag{3.6}$$

A priori, it is not obvious (especially for $p > 3$) that the ratio of displacement of the summit $\boldsymbol{s}_i$ along the segement $[\boldsymbol{v}_i, \boldsymbol{c}]$ translates into the same ratio of volume reduction for the opinion subdomain. The proof of this fact (see Appendix 3.6.3) relies on an appropriate decomposition of the opinion space into sub-simplices. We show that the sliding of the summit $\boldsymbol{s}_i$ towards $\boldsymbol{c}$ only scales the height of sub-simplices, without affecting their base, resulting in the same scaling in the volume of each sub-simplex. We will refer to the final opinion space (after ordering and reduction) as $C \subset S_{p-1}$.

3.1.2 Natural Opinions

The distribution of opinions underlying an election is largely impacted by the distribution of natural opinions. Similar to the natural opinion distribution of two-party system given by Eq. (2.6), we may consider polarization into more than two peaks, which naturally leads to case of multi-party systems. In that case, we did not consider polarization beyond the existence of parties and instead generated uniformly distributed natural opinions within the admissible opinion domain as described in the Appendix 3.6.3.

3.1.3 External Influence

The matrix $W^u \in \mathbb{R}^{n \times p}$, indexed by the electoral unit u, models the targeted influence of an external source on each agent. The equilibrium state of the system given by Eq. (3.3) remains on the simplex only if the opinions of the agents along with the external influence remain on the simplex, i.e. $((X^0)^u + W^u)_i \in S_{p-1} \, \forall \, i$. Otherwise, the final opinion $(X_\infty)^u_i$ of some agent i will have negative components, which takes the opinion of agent outside the simplex. According to our model, the natural opinion of each agent lies in the simplex, i.e. $(X^0)^u_i \in S_{p-1} \, \forall i$. Thus, the external influence applied on each agent i given by W^u_i should satisfy the following conditions:

1. $\sum_q ((X^0)^u + W^u)_{iq} = 1 \implies \sum_q W^u_{iq} = 0 \, \forall \, i$,

2. $0 \leq ((X^0)^u + W^u)_{iq} \leq 1 \implies -(X^0)^u_{iq} \leq W^u_{iq} \leq 1 - (X^0)^u_{iq}$.

Similar to the two-party system, we define a measure of *robustness* of an election outcome as the minimal amount of external influence required to change the outcome. We start with an uninfluenced system ($W_0 = \mathbf{0}_{n \times p}$) and add small increments of influence until the election outcome changes. We chose the influence increments to be as effective as possible, namely,

- influence increments are directed towards the first runner-up of the election, which has the most chance to steal the elections from the winner;

- agents whose natural opinion is close to the barycenter $\boldsymbol{c}$ are targeted first as their vote will arguably be the first to change;

- when multiple electoral units are involved, we target first the electoral units with the lowest majority as less agents needs to be targeted to change the winner in that electoral unit (verified to be the optimal strategy in two-party system in Chap. 2).

In summary, when we want to target the i^{th} agent, and push their opinion towards the q^{th} party (first runner-up of the election), we place a vector

$$\boldsymbol{w} = (w_1, \cdots, w_p) \perp \mathbf{1}_p \,, \tag{3.7}$$

to the i^{th} row of external influence matrix W^u. The components satisfy $\sum_i w_i = 0$ with $w_q > 0$ and $w_{q'} < 0$ for $q' \neq q$. For our numerical simulation, the values of w_q and $w_{q'}$ are arbitrarily chosen following the left-right political ordering similar to the natural opinions of agents such that we set $\sum_i |w_i| = 0.1$.

If the outcome of the election in electoral unit u changes after K_u increments (i.e. after influencing K_u agents), we define the *effort* needed to change the election outcome as the percentage of agents influenced,

$$\xi_u = \frac{K_u}{n_u} * 100 \,, \tag{3.8}$$

where the subscript u indexes the electoral units. When multiple electoral units are involved, the total effort is

$$\xi = \sum_u \frac{n_u}{N} \xi_u \,, \tag{3.9}$$

where N is the total number of agents in the whole country.

3.2 Result of Elections

A multi-party electoral system allows different ways of electing the winner as discussed in Chap. 1. For a system with a single electoral unit, we mainly focus on two different types of electoral systems as mentioned below.

Plurality System: In a plurality system the party that receives the most votes is declared the winner. This means that a party can win even if they did not secure more than 50% of votes. It is a simple electoral system widely used in various countries worldwide, for example Canada and India.

Ranked Choice Voting System (RCV): The agents in the electoral system are given the choice to rank their preferences. If a party gathers more than 50% of the first-priority votes, they are declared the winner. If no party wins after counting the first-priority votes, then the party with least number of first-priority votes gets eliminated. The corresponding second priority votes are counted to determine the winner. This process goes on, until one party gets more than 50% of votes, and we declare this party as the winner. The RCV system is commonly used in Ireland, Australia and so on.

In a system with multiple electoral units, there are multiple ways to select one or more representatives from each electoral units based on the number of votes that parties gain. Here we consider four main ones:

Proportional representation (PR): A set of m seats in an electoral unit are distributed among the parties to match their proportion of votes as closely as possible.

Single representative (SR): The electoral unit elects the candidate with majority of votes as its unique representative.

Winner-Takes-All (WTA): All seats associated within an electoral unit are attributed to the party with the most votes. If there is a single seat in the electoral unit, WTA reduces to SR.

Proportional Ranked Choice voting system (PRCV): Agents in each unit rank the candidates in order of preference. All candidates that receive a voting ratio above a certain quota are elected. Usually the quota is fixed around $1/(m+1)$ with the number m of seats in the considered electoral unit. Next, surplus votes above quota are re-distributed to not-yet-elected second-choice candidates. If new candidates exceed the quota, they are elected and their surplus votes are again re-distributed. If no candidate exceeds the quota, the candidate with the least vote is eliminated and their votes redistributed according to second preferences. The process is iterated until m seats are attributed.

3.3 Single Electoral Unit

In this section, we evaluate different types of electoral systems in a single electoral unit. Our first goal is to determine the winner and the first runner-up based on the party's position on the left-right political spectrum as a function of ϵ. Second, we analyze the effort required to alter the election outcome in favor of the first runner-up. Such an analysis provides insights into the level of influence required to shift the support towards other parties to change the election outcome. For our analysis, we consider a single electoral unit consisting of 2001 agents evenly distributed among the parties. We then compute the percentage of wins of each party over multiple realizations of natural opinions as a function of the confidence bound ϵ. We divide this section into two subsections on plurality and ranked choice voting systems.

3.3.1 Plurality System

In general, the system's dynamics depends on the interaction between the agents and their positions in the left-right spectrum. It is intuitive that if an agent i who supports party q interacts with many agents supporting party q', then the agent i tends to change its opinion and vote for party q'. Fig. 3.2 shows the percentage of wins and the percentage of first runner-ups along with the effort needed to change the elections outcome in a single electoral unit with six parties as a function of the confidence bound ϵ. Similar results of other parties (three, four, five and seven) are presented in Fig. 3.3.

Our models are symmetric in that the percentage of wins, of second, third and so on place finish are symmetrically distributed over parties ordered from 1 (leftmost) to 6 (rightmost) as shown in Fig. 3.2 **a**. When $\epsilon = 0$, the agents do not interact with each other and vote according to their natural opinion. Thus all parties have an equal percentage of wins irrespective of their position in the left-right political spectrum. Similar results holds for very small values of ϵ, due to very few interactions among the agents (see Figs. 3.2 and 3.3). However, as ϵ increases, percentage of wins of the parties at the center increases as the centrist parties establish more interactions with other agents compared to the moderates and extremists parties. In addition the attractive dynamics drives the agents towards the barycenter c. Thus the percentage of wins of the extremists in the election vanishes. With a well-connected network, i.e. for $\epsilon \gtrsim 0.4$, the centrists wins the election, and the party adjacent to the centrist becomes the first runner-up. If p is even, then the percentage of wins of the election is equally shared between the two centrists, and the same applies to the first runner-up.

Fig. 3.2 **b**, shows the effort needed to change the election outcome as a function of confidence bound ϵ. To understand the behavior of the effort needed to change the election outcome in favor of the first runner-up as a function of ϵ, we divide the total effort (black solid curve in Fig. 3.2 **b**) into its components based on the first runner-up's position in the left-right spectrum. In Fig. 3.2 for six party system, we can have extremists (party 1 or 6), moderates (party 2 or 5) and centrists (party 3 or 4) as the first runner-ups. Similar grouping of parties can be done for any system with p parties. In case of odd number of parties, there will be only one centrist party.

From Fig. 3.2 **b**, we observe that the effort needed to change the election outcomes increases with an increase in ϵ, reaches a maximum value at ϵ_m and then decreases with further increase in ϵ. One sees that, unlike the two-party case, the effort is here non-monotonous in ϵ. This is however an artifact of having extremist parties connected to few neighboring parties when ϵ is small, as we next proceed to argue. When ϵ is small, there are not many changes in opinion of agents due to few interactions among the agents, thus extremists have a chance to become first runner-ups.

When ϵ is large enough, agents start interacting with agents of neighboring parties and with the attractive dynamics of the model, agents tend to change their opinions towards the barycenter opinion. The agents of one party can convince the agents belonging to the adjacent party more efficiently compared to the agents belonging to other parties that are farther away. Thus, the extremists can convince only the agents from one adjacent party compared to other

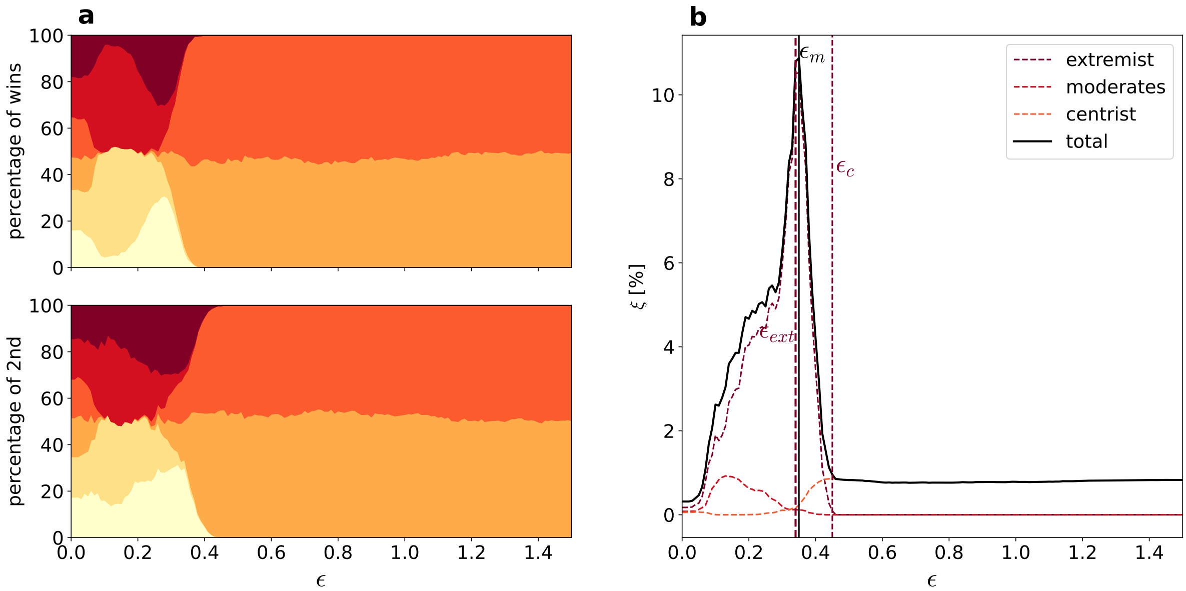


Figure 3.2: Robustness of electoral systems with six parties. **a**: Percentage of wins (top panel) and of second place (bottom panel) for each party in the absence of external influence. Different colors correspond to different parties in a left-right order when going from light yellow to dark red. Larger interaction distances favor centrists parties. **b**: Effort needed to overturn a 6-party election under the single representative electoral system in a single electoral unit (solid black curve). The total effort is broken down into partial efforts corresponding to overturns in favor of a centrist (dashed yellow), a moderate (dashed orange) or an extremist (dashed red) party. The total effort is dominated by the effort devoted to overturns in favor of an extremist.

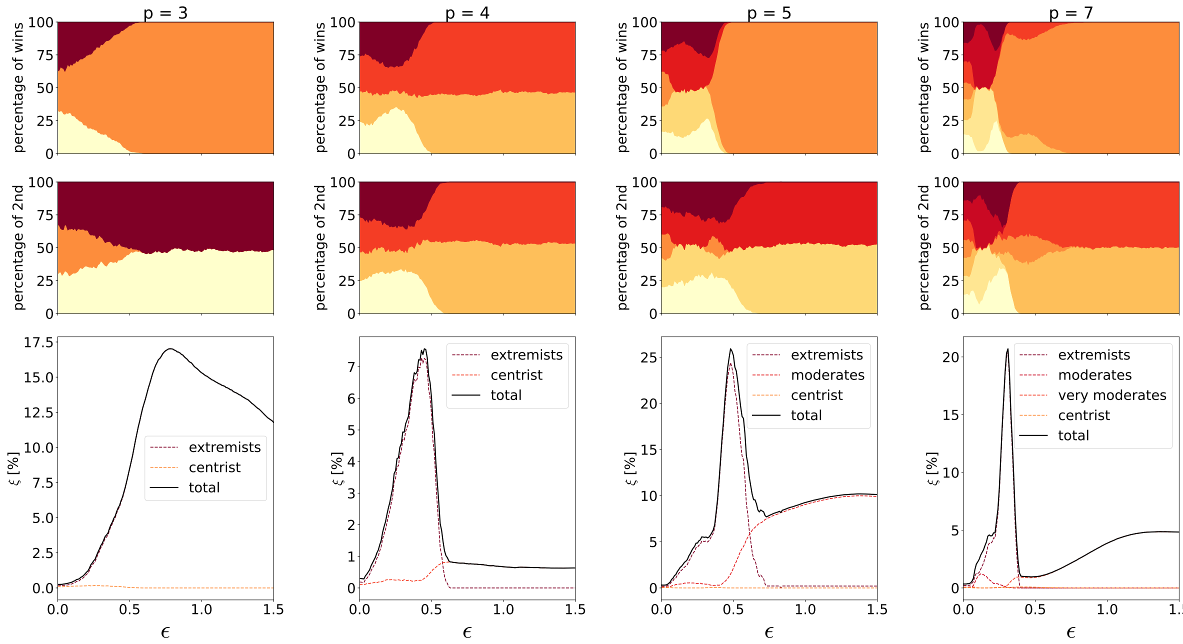


Figure 3.3: Same as Fig. 3.2 panels **a** and b, for three, four, five, and seven parties (from left to right) for plurality system. Top row: Percentage of times each party wins the elections. Middle row: Percentage of times each party is the first runner-up. Bottom row: average effort to change the election outcome, as a function of the confidence bound ϵ. The effort is broken down into the contribution of different realizations aggregated according to the position of the first runner in the political spectrum.

parties with two adjacent neighbors, due to its position in the left-right spectrum (see Figs. 3.2 and 3.3) and the extremists get attracted towards the centrists. As a result, the centrists become the winner or first runner-up, and the effort needed to change the election outcome is comparatively smaller compared to moderates or extremists. Thus the effort needed to change the election outcome decreases with large values of ϵ.

From Fig. 3.3, we observe that the confidence bound ϵ at which the effort is maximum ϵ_m changes with the number of parties. This change in ϵ_m is significant from a three to four-party system. For smaller values of ϵ, the probability of an extremist winning or being the first runner-up is the same compared to all the other parties (see Figs. 3.2 and 3.3). However, with an increase in ϵ, probability of extremists becoming the winner or first runner-up decreases and vanishes. Let ϵ_c be the largest ϵ at which at least one of the extremists manages to be the first runner-up, implying that after ϵ_c the contribution to the effort needed to change the election outcome from the extremists is zero. Furthermore, let ϵ_{ext} be the value of confidence bound ϵ at which the contribution to the total effort from extremists being the first runner-up is the maximum.

For a three-party system, we don't have a ϵ_c as the extremist never vanishes from being first runner-up, and the total effort needed to change the election outcome depends on influencing the system in favor of the extremist for all values of ϵ. As the effort needed to change the election outcome in favor of extremists is much higher than moderates and centrists, the ϵ at which we obtain the maximum effort needed to change the election outcome tends to lie in the interval $[0, \epsilon_c]$. For $\epsilon < \epsilon_c$, all the parties contribute to the effort needed to change the election outcome, and the significant contribution comes from the extremist. Thus, ϵ_m tends to depend mainly on the contribution of effort needed to make the extremists the winner. Since the contribution of the effort needed to make the other parties the winner are comparatively much smaller, we expect that ϵ_m to be almost the same as ϵ_{ext}. This line of reasoning is confirmed numerically as shown in the Fig. 3.2.

We could also observe that the ϵ_c becomes slightly smaller with an increase in the number of parties. The increase in the number of parties leads to more parties in the center that can interact more effectively than the extremists. Consequently the extremist loses its impact more easily than when there are fewer parties between them. According to our model, it is very hard for extremists to influence enough agents to take first place in an election. In addition, the farther the first runner-up is from the center of the left-right spectrum, the harder it becomes to influence enough agents to take first place in the election.

3.3.2 Ranked Choice Voting System

In this section, we analyze the behavior of the elections subjected to the Ranked Choice Voting System (RCV). As mentioned earlier, the system's dynamics remain the same but the election differs in how the votes are counted. The system's dynamics depend on the interaction between the agents, attractive dynamics, and natural opinions. In addition, for RCV system the second-priority votes also have a significant role in determining the winner of election.

In Figs. 3.4 and 3.6 we present the analysis of percentage of wins and the effort needed to change the election outcome in favor of first runner-ups using the RCV system. The percentage of wins of each party depends on the confidence bound ϵ. Additionally, we observe a non-monotonous behavior for the percentage of wins for odd-party systems as a function of ϵ. For large values of ϵ the percentage of wins of the election is shared between two parties at the center of left-right political spectrum for an even number of parties and between three parties at the center for an odd number of parties. The percentage of wins as a function of ϵ is intuitive for three and four party systems. However, it becomes challenging when there are more number of parties and needs more detailed considerations.

In order to understand the non-monotonous behavior of the percentage of wins corresponding to odd-party systems as a function of ϵ, we consider the case of five-party systems (see Fig. 3.4). Fig. 3.4 **a** shows the percentage of wins of each party, in Fig. 3.4 **b**(top panel) we show the average percentage of votes of each party, and in Fig. 3.4 **b**(bottom panel) the percentage of 5th position, i.e. the party that would get eliminated in the first round. Finally in Fig. 3.4 **c**, we show the percentage of 2nd priority votes of extremist, moderates and centrist. Here we show the right-wing parties, same corresponding figure hold for left-wing due to the symmetry of our model. For ease of explanation, we divide the ϵ-range into four different intervals $[0, 0.225]$, $[0.225, 0.4]$, $[0.4, 0.9]$, and $[0.9, 1.5]$. These intervals correspond to values of ϵ with approximately the same trend; either having an increase or a decrease in the percentage of wins for the centrists as a function of confidence bound ϵ.

For smaller values of $\epsilon \in [0, 0.225]$, the percentage of wins of the centrist decreases as ϵ increases. The percentage of centrists receiving the least first-priority votes increases in this range as shown in Fig. 3.4 **b**. As a result, the centrist gets eliminated in the first elimination step by distributing their votes almost equally to both the moderates as shown in Fig. 3.4 **c**. Thus the percentage of wins of centrist decreases for $\epsilon \in [0, 0.225]$. With the new votes received

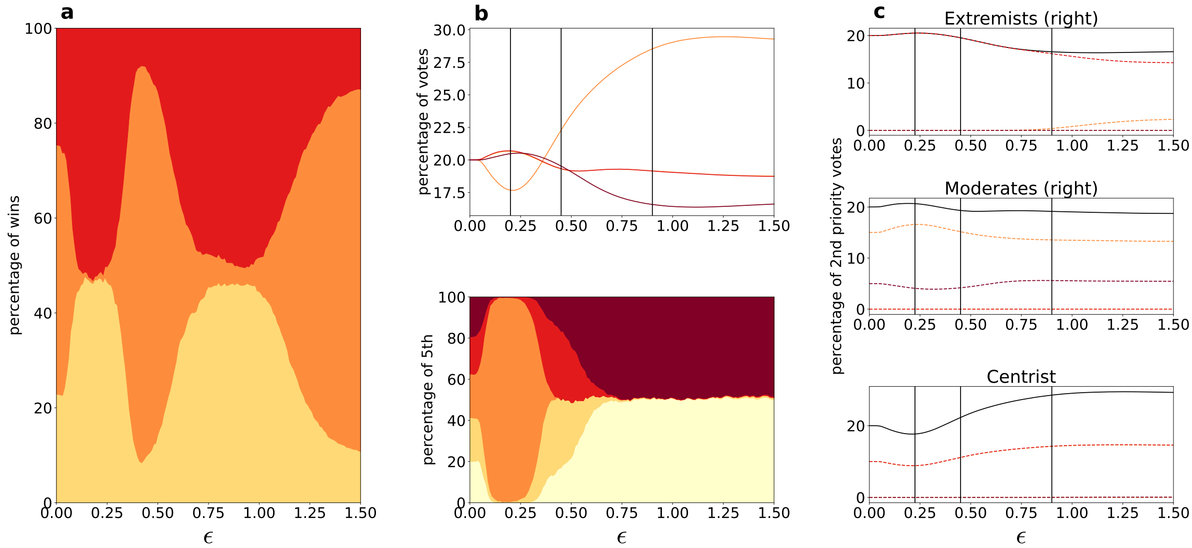


Figure 3.4: Percentage of wins analysis for five party system. **a**: Proportion of realizations where each party is the winner. **b**: The average percentage of votes per party over the realization (top panel) and proportion of realization where each party is 4th runner-up. **c**: The distribution of second priority votes of left extremist(top panel), left moderates (middle panel) and centrists(bottom panel) to other parties. The color scale denotes the ordering of the parties, from light yellow for the left-most party to dark red for the right-most one.

from the centrists, the percentage of votes acquired by the moderates increases, whereas there is no change in the percentage of votes of extremists. This results in the elimination of extremists in the next elimination step. During the elimination of extremists, the moderates next to the eliminated extremist gains more votes and wins the elections. This results in an increase in the percentage of wins for the moderates.

For confidence bound $\epsilon \in [0.225, 0.45]$, the average percentage of first-priority votes of centrists increases, and the corresponding percentage of votes of the moderates and extremists decreases (see Fig. 3.4 **b** (top panel)). As a result, the centrists decrease the probability of getting eliminated in the first elimination step, and mostly, the moderates get eliminated (see Fig. 3.4 **b** (bottom panel)). When the moderates get eliminated, their votes are redistributed among the centrists and extremists, with a major share to the centrists than the extremists, thus the percentage of votes of the centrists increases (see Fig. 3.4 **c** (middle panel)). This decreases the chance of centrists getting eliminated in the next elimination step. With the votes acquired from moderates the probability of centrists getting eliminated in the subsequent elimination vanishes, and the percentage of wins of centrists increases.

For confidence bound $\epsilon > 0.4$, the centrists acquire more first-priority votes than the other parties because of the attractive dynamics and their position in the left-right spectrum of political parties, and the extremists have the least votes. For moderate values of $\epsilon \in [0.4, 0.9]$, all the votes of extremists after their elimination get redistributed to the moderates (see Fig. 3.4 **c** (top panel)). The percentage of votes of the moderates increases above the percentage of votes of centrists, and the centrist gets eliminated in the subsequent rounds before the moderates. Thus, the percentage of wins of centrists decreases again.

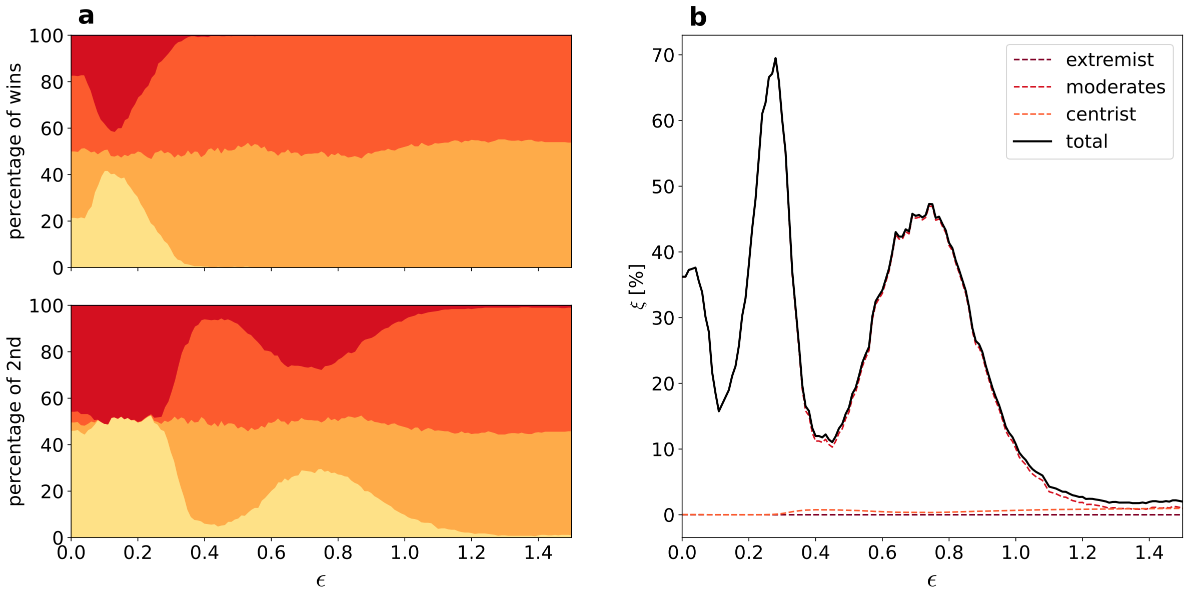


Figure 3.5: Robustness of electoral systems with six parties using RCV system. **a**: Proportion of realizations where each party is the winner (top panel) and the first runner (bottom panel). The colour scale denotes the ordering of the parties, from light yellow for the left-most party to dark red for the right-most one. One notices extremists can never get a decent score (winner or first runner) for any value of communication distance. **b**: Breakdown of the effort curve (plain black) for a single electoral unit with six parties. The dashed lines show the contribution to the black curve of the effort dedicated to the extremist parties (dark red), the moderates (orange), and the centrists (yellow). The extremists have no contribution to the curve, and the moderates have a large contribution to the curve (see panel **b**).

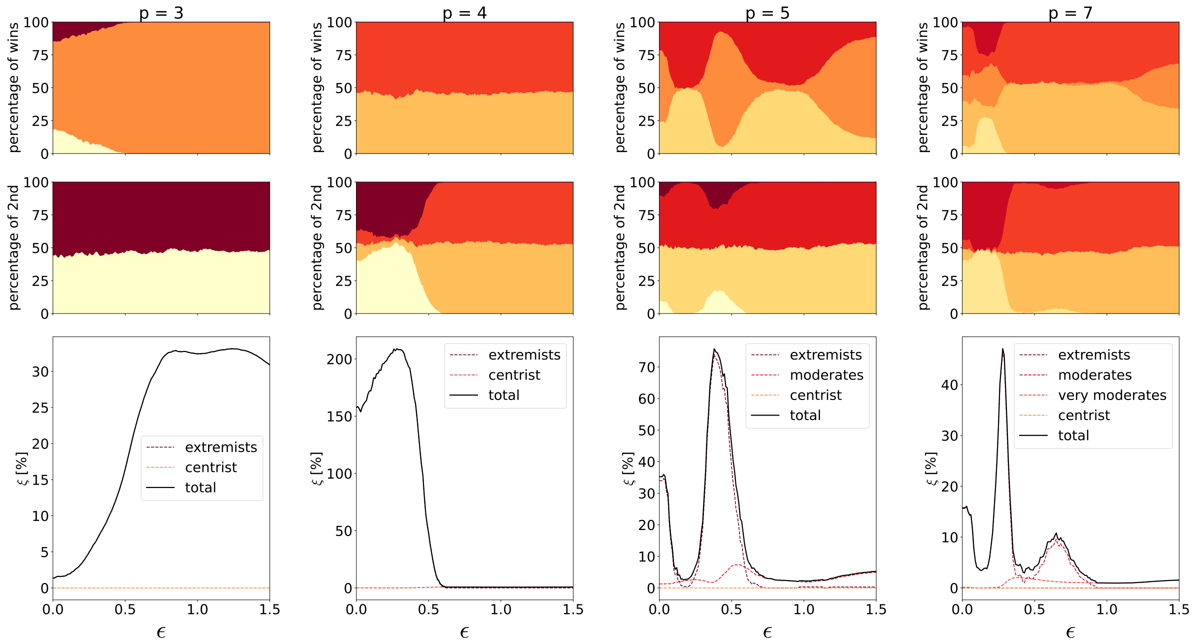


Figure 3.6: Same as Fig. 3.4 panels a and b in the main text, for three, four, five, and seven parties (from left to right) for ranked choice voting. Top row: Percentage of times each party wins the elections. Middle row: Percentage of times each party is the first runner-up. Bottom row: average effort to change the election outcome, as a function of the confidence bound ϵ. The effort is broken down into the contribution of different realizations aggregated according to the position of the first runner in the political spectrum.

For larger values of $\epsilon \in [0.9, 1.5]$, the extremists will still get eliminated in the first round. However, the system's dynamics do not conserve the ordering of the parties for large values of ϵ, and the votes of extremists are distributed to centrists and moderates. As a result, the percentage of votes of centrists and moderates increases. In the next round of elimination, the other extremist will gets eliminated by distributing their votes. Thus, the centrist gain votes from both the extremists, whereas the moderates gain votes only from the extremists next to them, and the percentage of wins for centrists increases. This explains the non-monotonous behavior of the percentage of wins as the function of ϵ.

Fig. 3.5 shows the effort needed to change the election outcome in a single electoral unit with six parties, averaged over independent realizations of natural opinion (black solid line). Similar plots are presented in Fig. 3.6 for other parties. The colored dashed lines show the breakdown of the overall average effort into the different possibilities for the first runner-up, distinguished by how far the party is from the center of the political spectrum as discussed in Sec. 3.3.1. We see that the shape of the effort curve is mostly determined by the realizations where one of the moderates is the first runner-up. In the RCV system, the moderates play the significant role in determining the shape of the effort curve. This is due to the fact that, among the first runner-ups, the moderates are farther away than the centrists from the center of the left-right political spectrum. As observed from the Fig. 3.5, the non-monotonous behavior of the effort follows from the non-monotonous behavior of the first runner-ups as shown in Fig. 3.5 **a** (bottom panel). According to our model, we conclude that the farther the first runner-ups are from the center of the political spectrum, the harder it is to influence enough agents to win the election, similar to the plurality system.

3.4 Multiple Electoral Units

In this section, we evaluate the robustness of different types of electoral systems in a multi-party system of elections with multiple electoral units. We compare the robustness of the four electoral systems (SR, WTA, PR, PRCV), as discussed in Sec. 3.2 for three to seven parties. To evaluate the robustness of the electoral system we consider only the effort needed to change the election outcome for the first runner-up, irrespective of the number of parties participating in the election.

We consider the same set of synthetic countries considered for two-party system as mentioned in Sec. 2.4.2. For numerical simulations, we consider 101 agents corresponding to each seat in an electoral unit, as per the defined synthetic country. The electoral unit for simulation is the state, and as mentioned in Chap.2, the number of seats per state is randomly drawn from the interval $I = [3, 15]$, and the interaction network is constructed at the state level. For each realization, a random difference of votes is introduced with the maximum difference of votes between any two parties being 10%. In each electoral unit, we tune the distributions of natural opinions to agree with the percentage of votes corresponding to each party. Finally, we average the effort of over 15 different synthetic countries and realizations of natural opinions for each of the different electoral systems.

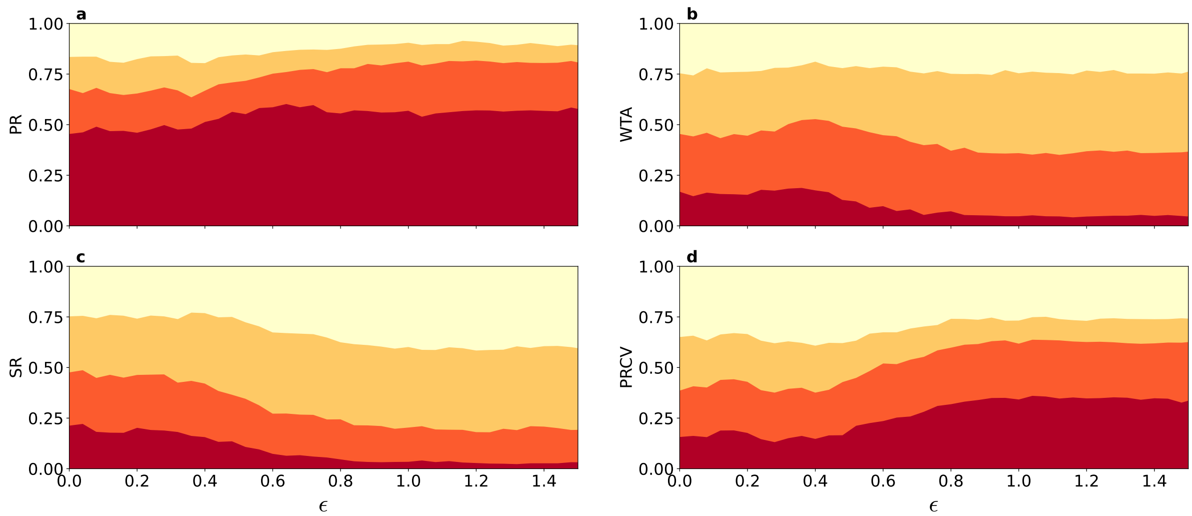


Figure 3.7: Robustness of different electoral processes for a six-party system vs. confidence bound over random synthetic countries with multiple electoral units. For each realization, a random difference of votes (uniform between 0% and 10%) is introduced. The plots show the proportion of realizations where each electoral system was ranked first (dark red), second (orange), third (yellow) or fourth (light yellow) in terms of robustness for **a**: PR, **b**: WTA, **c**: SR, **d**: PRCV.

Fig. 3.7 shows the robustness of electoral systems as a function of confidence bound ϵ for six party system. Here we consider the proportion of realizations where each electoral system

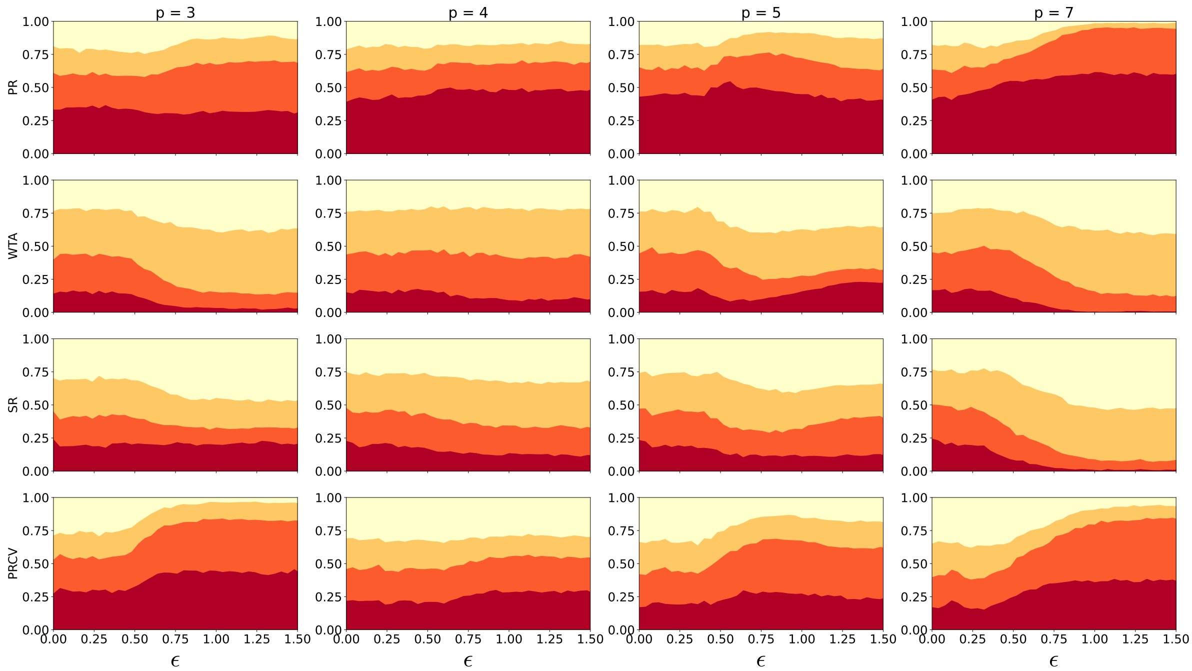


Figure 3.8: Same as Fig. 3.7, for three, four, five, and seven parties (from left to right) and for the four electoral systems: PR, WTA, SR, and PRCV (from top to bottom). Proportion of realizations where the electoral system is the most robust (dark red), second most robust (orange),third most robust (yellow), and least robust (light yellow) in our simulations.

was ranked first (dark red), second (orange), third (yellow) or fourth (light yellow) in terms of robustness. On evaluating the robustness of the different electoral systems, it clearly shows that the PR system is the most robust, most of the time, with a percentage of first-rank robustness starting slightly below 50% at small ϵ and increasing to about 65% with increasing ϵ. The second most robust system is the PRCV system, if judged from first-ranked robustness. Note however that it is the least robust about 30% of the time. As for the two-party case, the WTA system is more robust than the SR system. The SR is generally the least robust to external influence for multi-party systems.

Fig. 3.8 shows the robustness of electoral systems as a function of confidence bound ϵ for other parties (three, four, five and seven). The robustness of electoral system doesn't have much effect on the number of parties, and the results remains similar to the six-party system with PR being the most and SR being the least robust electoral system. Our results are also consistent with the two-party system as PR and SR system are generally the most and least robust electoral system respectively.

3.5 Conclusion

We extended our model to a multi-party system of elections as most countries have more than two major parties competing in the elections. Following the same considerations as in the two-party electoral system, we extend our model to accommodate any number of parties along with the inherent left-right political spectrum in the society. Here we evaluate two most common multi-party electoral systems: plurality and ranked choice voting systems.

Unlike the two-party electoral system, the effort needed to change the election outcome as a function of confidence bound ϵ is non-monotonous. The major contribution to the total effort is from the first runner-up party which is farther away from the centrists. The effort needed to change the election outcome depends on how far the first runner-up is from the center of the left-right spectrum. i.e. it is more difficult for the extremists to convince the agents to vote for them compared to any other parties. However this is an artifact of having extremists having just one neighboring party compared to other parties which have two neighboring parties.

In the ranked choice voting system, the percentage of wins depends on the agents' interactions and the agents' second-priority votes belonging to the eliminated party. Similar to the plurality system, the effort curve of the RCV system is determined by the position of the first runner-up in the left-right spectrum. Similar to the plurality system, the effort follows the contribution from the first runner-up party, that is farther away from the centrists.

On evaluating different types of electoral systems in multiple electoral units, we observe that the PR is the most robust, followed by the PRCV, WTAR and SR electoral systems. We can conclude that the most and least robust electoral systems are PR and SR respectively.

3.6 Appendix

3.6.1 Equivalence to One-Dimensional Dynamics

In this section, we verify that the model for a multiparty system given by Eq. 3.1 is equivalent to the previous model in one-dimensional for two parties discussed in Chapter 2. Let $\boldsymbol{x}^0 \in [-1,1]^n$ be the natural opinion corresponding to the previous model. We omit the dependence on electoral unit 'u' for ease of notation. Now we define a projection $X : [-1,1]^n \to (S_1)^n$ such that

$$(X(\boldsymbol{x}^0))_i = [0.5 - 0.5x_i^0, 0.5x_i^0 + 0.5]. \tag{3.10}$$

To prove that the two models are equivalent, it is enough to prove that Y is a bijection and the dynamics are equivalent. It is trivial to see that Y is a bijection. For every $y_1 \in (S_1)^n$, there exist only one $x_i \in \mathbf{x^0}$ such that $X_1(x_i) = 0.5 - 0.5x_i$ and the statement hold for X_2 also. Thus Y is a bijection. Now to prove that the dynamics are equivalent, it is enough to prove that $\dot{X} = \frac{d}{dt}X(x)$. This can be verified by direct computation.

$$\begin{aligned} \dot{X}_i &= \frac{d}{dt}X(x_i) \\ &= \frac{d}{dt}\Big[0.5 - 0.5x_i, 0.5x_i + 0.5\Big] = \Big[-0.5\dot{x_i}, 0.5\dot{x_i}\Big] \end{aligned} \tag{3.11}$$

Substituting Eq. (2.3) into Eq. (3.11), and grouping them we get

$$\dot{X}_i = -\sum_j (D_\epsilon^{-1}L_\epsilon + \mathbb{I})_{ij}[0.5 - 0.5x_j, 0.5 + 0.5x_j] + \Big[0.5 - 0.5y_i^0, 0.5 + 0.5y_i^0\Big], \tag{3.12}$$

where $y_i^0 = x_i^0 + \omega_i$. This is rewritten as

$$\dot{X}_i = -\sum_j (D_\epsilon^{-1}L_\epsilon + \mathbb{I})_{ij}X_j + (X_i^0 + W_i), \tag{3.13}$$

following Eq. (3.10), which proves that the general model for two party is equivalent to the previous model.

Proposition 8. *If $X^0 \in (S_{p-1})^n$, then $X(t) \in (S_{p-1})^n$ $\forall t$.*

Proof. First, by definition, $(X^0 + W)_{ij} \geq 0$ for all $i \in \{1, \cdots, n\}$ and $j \in \{1, \cdots, p\}$. First we will prove that $X_\infty \in (S_{p-1})^n$. According to Eq. (3.3),

$$X_\infty = \Big(D_\epsilon^{-1}L_\epsilon + \mathbb{I}\Big)^{-1}\Big(X^0 + W\Big) = M^{-1}(X^0 + W). \tag{3.14}$$

From Prop. 6, M^{-1} is row-stochastic. First we prove that,

$$(X_\infty)_{ij} \geq 0\,, \tag{3.15}$$

for all $i \in \{1, \cdots, n\}$ and $j \in \{1, \cdots, p\}$. By definition,

$$\begin{aligned} 0 \leq &\quad (X^0 + W)_{kj} &&\leq 1, \\ 0 \leq &\quad (M^{-1})_{ik} \cdot (X^0 + W)_{kj} &&\leq (M^{-1})_{ik}, \\ 0 \leq &\quad \textstyle\sum_{k=1}^n (M^{-1})_{ik}(X^0 + W)_{kj} &&\leq \sum_{k=1}^n (M^{-1})_{ik}, \\ 0 \leq &\quad (X_\infty)_{ij} &&\leq 1. \end{aligned}$$

Second, let us then compute

$$\begin{aligned}\sum_j (X_\infty)_{ij} &= \sum_j \left(M^{-1}X^0\right)_{ij} \\ &= \sum_j \sum_k \left(M^{-1}\right)_{ik} X^0_{kj} \\ &= \sum_k \left(M^{-1}\right)_{ik} \sum_j X^0_{kj} \\ &= \sum_j \left(M^{-1}\right)_{ik} \\ &= \left(M^{-1}\mathbf{1}\right)_i = 1\,,\end{aligned} \tag{3.16}$$

for all $i \in \{1, \cdots, n\}$, where we used that $X^0_i \in S_{p-1}$ from third to fourth line.

Eqs. (3.15) and (3.16) are exactly the conditions defining the S_{p-1}, thus $X_\infty \in (S_{p-1})^n$.

Solving Eq. (3.2) yields

$$X(t) = e^{-Mt}X(0) + \left(\mathbb{I} - e^{-Mt}\right) \underbrace{M^{-1}\left(X^0 + \boldsymbol{\omega}\right)}_{Z} \tag{3.17}$$

We know that the final point of the opinions trajectory is in the simplex. Now in order to prove that the whole trajectory remain in the simplex. We check the first condition for $X_i(t)$ to be in the simplex, i.e.,

$$\sum_j X_{ij}(t) = 1\,, \qquad\qquad \forall\, t \geq 0\,. \tag{3.18}$$

By direct calculation,

$$\begin{aligned} X(t)\mathbf{1} &= (e^{-Mt}X(0) + \left(\mathbb{I} - e^{-Mt}\right) Z)\mathbf{1} \\ &= e^{-Mt}X(0)\mathbf{1} + \left(\mathbb{I} - e^{-Mt}\right) Z\mathbf{1} \\ &= e^{-Mt}\mathbf{1} + \left(\mathbb{I} - e^{-Mt}\right)\mathbf{1} \\ &= \mathbf{1}\end{aligned} \tag{3.19}$$

It is to be noted that we employed the fact that Z is row-stochastic, which follows from the product of two row-stochastic matrix is row-stochastic.

Now consider the linear consensus dynamics with influence vector given by (3.2) in component wise

$$\begin{aligned} \dot{X}_i &= -D^{-1}_{ii} \sum_k (L_\epsilon)_{ik} X_k - (X_i - X^0_i - \omega_i) \\ \dot{X}_{ij} &= -D^{-1}_{ii} \sum_k (L_\epsilon)_{ik} X_{kj} - (X_{ij} - X^0_{ij} - \omega_{ij}) \\ \sum_j \dot{X}_{ij} &= -D^{-1}_{ii} \sum_k \sum_j (L_\epsilon)_{ik} X_{kj} - \sum_j (X_{ij} - X^0_{ij} - \omega_{ij}) \\ &= -D^{-1}_{ii} \sum_k (L_\epsilon)_{ik} \\ &= 0 \end{aligned}$$

Thus the dynamics of the system remain in the simplex. □

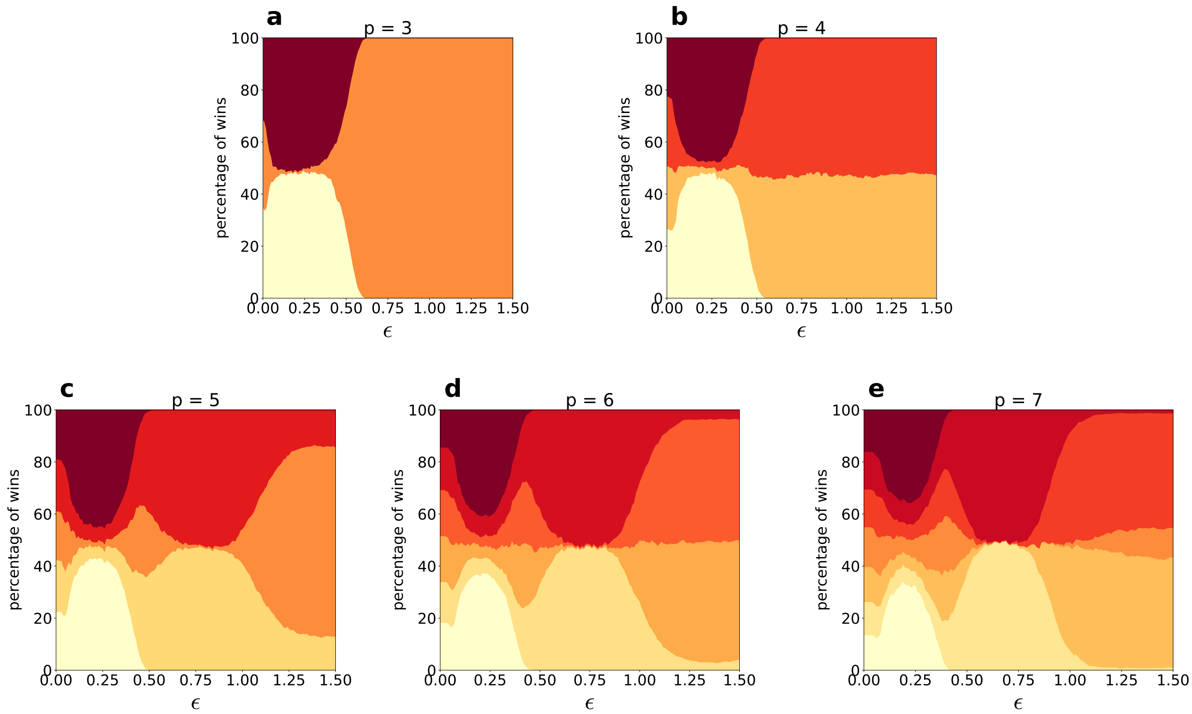


Figure 3.9: The figure shows the percentage of wins as a function of confidence bound ϵ for three, four, five, six and seven parties over 500 realizations of natural opinion. Different colors correspond to different parties in a left-right order when going from light yellow to dark red.

3.6.2 Effect of Ordering

In this section, we consider a single electoral unit of 2001 agents with p parties, similar to the first round of the French Presidential election. We aim to explore the effect of ordering in the model. For natural opinion without ordering, the probability that an agent in party i interacts with an agent in party j is the same for any party i and j. Whereas after ordering, the probability that an agent in party i interacts with an agent in party j depends on how aligned are their opinions. In simple words, ordering reduces the interaction between extremists because of their differing opinions.

In order to study the effect of ordering, we consider each party with almost equal strength; the percentage of wins of each party is the same when there is no interaction ($\epsilon = 0$). The interaction network is then constructed from the confidence bound ϵ. When there is no ordering, with an increase in confidence bound ϵ the interaction between all the agents increases and is independent of the party's position in the left-right political spectrum. Consequently, percentage of wins of each party also remains the same for all parties.

When the parties are ordered, even if the parties have almost equal agents supporting them the extent to which each party interacts within and outside varies for different parties. Since the ordering is symmetric around the centrist, the symmetry is maintained with an increase in confidence bound ϵ as well. It is interesting to observe that when the parties are ordered, there exists a range of confidence bound ϵ for which the percentage of wins of each party with zero influence is higher than the other parties. Thus percentage of wins exhibits non-monotonous behavior. Fig. 3.9 shows that the extremists wins for a smaller value of ϵ whereas the centrists wins for higher values of ϵ. We want to understand why ordering makes the extremist parties more likely to win the election for small values of ϵ and centrist parties are more likely to win for large ϵ. Let X^0_{ij} be the largest component of agent i. Then by ordering we have,

$$\begin{aligned} &X^0_{i1} < X^0_{i2} < \cdots < X^0_{ij}, \\ &X^0_{ip} < X^0_{i(p-1)} < \cdots X^0_{ij} \end{aligned} \tag{3.20}$$

Since agents i and j interact if $||X^0_i - X^0_j|| < \epsilon$, we define the neighbors of agent i as

$$N(i) = \{j \in \{1, 2, \cdots n\} \mid ||X^0_i - X^0_j|| < \epsilon\} \tag{3.21}$$

The final opinions of agents are given by

$$X_\infty = (D^{-1}_\epsilon L_\epsilon + \mathbb{I})^{-1} X^0. \tag{3.22}$$

One can verify that this series converges as follows. By Gershgorin's Circles Theorem [161, Theorem 6.1.1], all eigenvalues of $\mathbb{I} + D^{-1}_\epsilon A_\epsilon$ have real part in the interval $[0, 2]$, and therefore, the eigenvalues of $D^{-1}_\epsilon A_\epsilon$ have real part in $[-1, 1]$. Furthermore, it is straightforward to see that $\lambda = 1$ is an eigenvalue of $D^{-1}_\epsilon A_\epsilon$, with associated eigenvector $\mathbf{1}_n$. The matrix $D^{-1}_\epsilon A_\epsilon$ being nonnegative, by the Perron-Frobenius Theorem [161, Theorem 8.3.1], $\lambda = 1$ is actually the spectral radius $\rho(D^{-1}_\epsilon A) = 1$ and the spectral radius of $D^{-1}_\epsilon A_\epsilon/2$ is then $1/2$. Now, by Gelfand's formula [161, Corollay 5.6.14], we know that, for any matrix norm, there exists $k > 0$ such that $\|(D^{-1}_\epsilon A_\epsilon/2)^k\| < 1$. Therefore, considering the norm of the Neumann series,

$$\begin{aligned} \left\| \sum_{j=0}^{\infty} (D^{-1}_\epsilon A_\epsilon/2)^j \right\| &\le \sum_{j=0}^{\infty} \|D^{-1}_\epsilon A_\epsilon/2\|^j \\ &= \Big(1 + \|D^{-1}_\epsilon A_\epsilon/2\| + \cdots + \|D^{-1}_\epsilon A_\epsilon/2\|^{k-1}\Big) \sum_{j=0}^{\infty} \|D^{-1}_\epsilon A_\epsilon/2\|^{kj} \\ &< \infty\,, \end{aligned}$$

i.e., the series converges and its limit is precisely $(\mathbb{I} - D^{-1}_\epsilon A_\epsilon/2)^{-1}$. The inverse of the matrix $(D^{-1}_\epsilon \mathbb{L} + \mathbb{I})$ can then be expanded as

$$\begin{aligned} (2\mathbb{I} - D^{-1}_\epsilon A_\epsilon)^{-1} &= \frac{1}{2}[\mathbb{I} - D^{-1}_\epsilon A_\epsilon/2]^{-1} \\ &= \sum_{k=0}^{\infty} \frac{1}{2^{k+1}} (D^{-1}_\epsilon A_\epsilon)^k\,. \end{aligned} \tag{3.23}$$

The first-order approximation of the outcome is then given by

$$X^{(1)}_\infty = \frac{1}{2}\left[\mathbb{I} + \frac{1}{2} D^{-1}_\epsilon A_\epsilon\right] X^0. \tag{3.24}$$

The first-order approximation of final opinion of agent i for a party j is given by

$$\begin{aligned} (X^{(1)}_\infty)_{ij} &= \frac{1}{2}\sum_k (\mathbb{I} + \frac{1}{2} D^{-1}_\epsilon A_\epsilon)_{ik} X^0_{kj} \\ &= \frac{1}{2}\sum_k (\mathbb{I}_{ik} + \frac{1}{2}(D^{-1}_\epsilon A_\epsilon)_{ik}) X^0_{kj} \\ &= \frac{1}{2}\Big(X^0_{ij} + \frac{1}{2}(D^{-1}_\epsilon)_{ii} \sum_{k \in N(i)} X^0_{kj}\Big). \end{aligned} \tag{3.25}$$

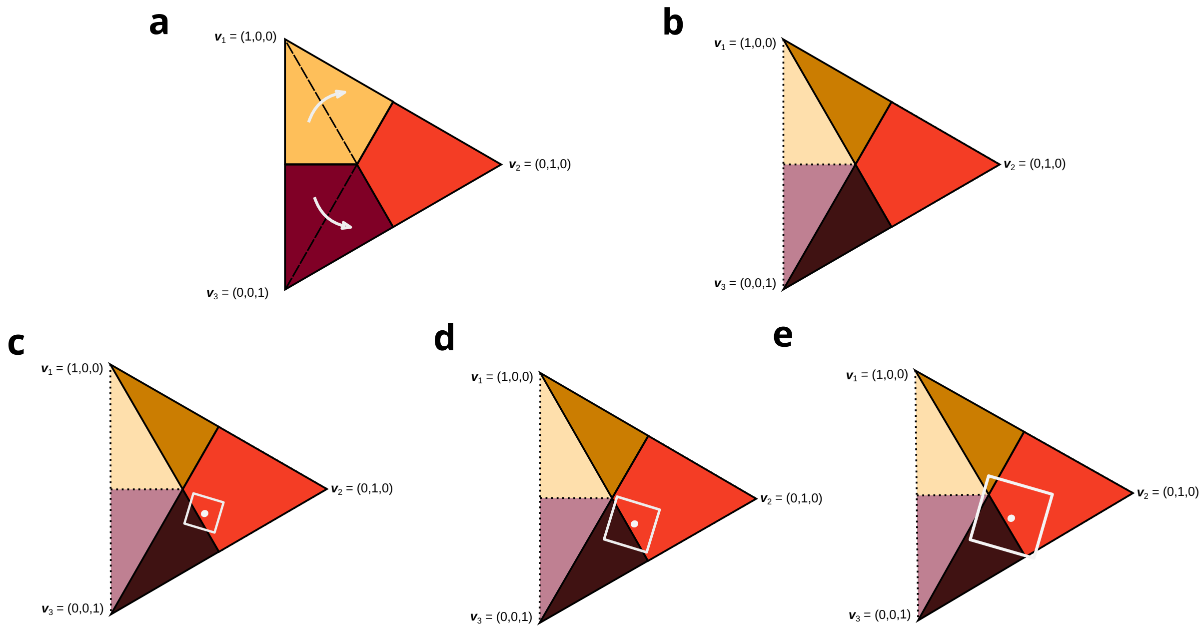


Figure 3.10: Illustration of the distribution of natural opinions over simplex for $p = 3$ parties [panels **a** and **b**]. Illustration of the area of confidence for a chosen agent and three values of communication distance [panels **c**, **d**, and **e**].

Now for $(X_\infty)_{ij}$ to be the largest component of agent i and votes for party j if

$$(X_\infty^{(1)})_{ij} > (X_\infty^{(1)})_{iq} \qquad \forall q \neq j,$$
$$X_{ij}^0 + \frac{1}{2}(D_\epsilon^{-1})_{ii} \sum_{k \in N(i)} X_{kj}^0 > X_{iq}^0 + \frac{1}{2}(D_\epsilon^{-1})_{ii} \sum_{k \in N(i)} X_{kq}^0 \qquad \forall q \neq j. \tag{3.26}$$

From Eq. (3.26), we see that the final opinion $(X_\infty)_i$ depends on its own opinion and of its neighbors.

For visual convenience, we consider the case of $p = 3$ parties. First of all, let us look at the distribution of opinions on the simplex before and after re-ordering. Fig. 3.10 **a** shows the density of opinion when we draw N random numbers uniformly in the 2-simplex. Due to the ordering of opinion, the areas in the two triangles along the edge between summits v_1 and v_3 are not allowed. The re-ordering process then transforms these opinions to their symmetric counterparts in the allowed region of the simplex, as illustrated by the arrows in Fig. 3.10 **a**, which yields the density illustrated in Fig. 3.10 **b**.

In order to understand how this distribution leads to a dominance of extremists for small ϵ and a centrist dominance for large ϵ, we will think about the continuous distribution on the simplex instead of a particular realization of natural opinions.

In the final state of first-order approximation, the opinion of an agent i is the mean of their own natural opinion and half of the average natural opinion of their neighbors, i.e., agents with natural opinion in an ϵ-radius of their own natural opinion. As illustrated in Figs. 3.10 [**c**-**e**], if an agent has a natural opinion very close to the boundary between two parties (red dot), almost twice as many of their neighbor will be partisan of the more extreme party. This is true at least for small values of ϵ [Fig. 3.10 **c**]. This remains more or less true until the boundary of the ϵ-ball around the natural opinion of agent i meets the boundary of the opinion domain [Fig. 3.10 **d**].

Then if ϵ is further increased, agent i will gain much fewer neighbors from the agents belonging to the extremists[Fig. 3.10 **e**]. Their final opinion will therefore be less driven towards extremist, as ϵ is so large that the interaction graph is complete and the process is averaging over all agents.

3.6.3 Opinion Volume Balancing

For a system with p parties, by definition [Eq. 3.4], the opinion space is a $(p-1)$-dimensional subset of $\mathbb{R}^p$. Therefore, for simplicity's sake, we embed the regular $(p-1)$-simplex S_{reg} in $\mathbb{R}^{p-1}$.

Let $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p \in \mathbb{R}^{p-1}$ be p points and let

$$S := \text{conv}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p)\,, \tag{3.27}$$

be the $(p-1)$-*simplex* $S \subset \mathbb{R}^{p-1}$ with *vertices* $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p$. The convex hull of $p-1$ vertices of S is a $(p-2)$-simplex, called a *facet* of S. The facet of S not containing the vertex $\boldsymbol{v}_i$ is denoted by F_i (it is the facet opposite to $\boldsymbol{v}_i$ in S), that is

$$F_i := \text{conv}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_{i+1}, \cdots, \boldsymbol{v}_p)\,, \qquad i \in \{1, \cdots, p\}\,. \tag{3.28}$$

The angle between the facets F_i and F_j is called the *dihedral angle* $\alpha_{ij} \in [0, \pi]$ of S.

The simplex $S \subset \mathbb{R}^{p-1}$ is said to be *regular* if

$$\alpha_{ij} = \alpha_{kl} =: \alpha_{\text{reg}} \forall 1 \leq i, j, k, l \leq p$$

It can be shown that $\alpha_{\text{reg}} = \arccos\frac{1}{p-1}$ [164, Section 7.9]. For $1 \leq q \leq p$, a $(q-1)$-simplex $S \subset \mathbb{R}^{p-1}$ is said to be a $(q-1)$-*orthoscheme* if $\alpha_{ij} = \frac{\pi}{2}$ if $|i-j| > 1$. For instance, a 2-orthoscheme is a right triangle, and the facets of a $(q-1)$-orthoscheme are $(q-2)$-orthoschemes.

The regular $(p-1)$-simplex S_{reg} can be dissected into $p!$ isometric $(p-1)$-orthoschemes by means of the recursive *barycentric decomposition* as follows (see Fig. 3.11). For $1 \leq k \leq p$, denote by $\boldsymbol{v}_{1,\cdots,k}$ the barycenter of the convex hull $\text{conv}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k)$. Because S_{reg} is regular (so that all its d-dimensional faces are regular d-simplices themselves, $0 \leq d \leq p-1$), $\boldsymbol{v}_{1,\cdots,k}$ is the center of the $(k-1)$-sphere inscribed in the convex hull of $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$ (and of the $(k-1)$-sphere circumscribed around the convex hull of $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$, etc.). For instance, $\boldsymbol{v}_{1,2}$ is the midpoint of the edge $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ of S_{reg}, $\boldsymbol{v}_{1,2,3}$ is the center of the regular triangle with vertices $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ (a 2-dimensional face of S_{reg}), and $\boldsymbol{v}_{1,\cdots,p}$ is the center of S_{reg}. Then, the convex hull

$$O_{1,\cdots,p} := \text{conv}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{1,\cdots,p})\,, \tag{3.29}$$

is a $(p-1)$-orthoscheme whose non-right dihedral angles are $\frac{\pi}{3}, \cdots, \frac{\pi}{3}, \frac{\alpha_{\text{reg}}}{2}$.

The procedure described above can be performed for any permutation of $\{1, \cdots, p\}$. Let $\mathcal{S}_p$ denote the group of all permutations of p objects, and let $(i_1, \cdots, i_p) \in \mathcal{S}_p$ be such a permutation. Then, the corresponding orthoscheme $O_{i_1,\cdots,i_p}$ is given by

$$O_{i_1,\cdots,i_p} = \text{conv}(\boldsymbol{v}_{i_1}, \boldsymbol{v}_{i_1,i_2}, \cdots, \boldsymbol{v}_{i_1,\cdots,i_p})\,. \tag{3.30}$$

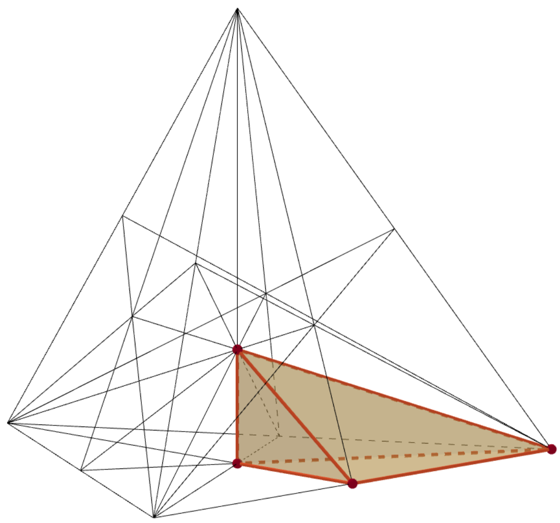


Figure 3.11: The regular 3-simplex S_{reg} and one of its 24 isometric fundamental orthoschemes O. Notice that here $p = 4$.

Because S_{reg} is regular, $O_{i_1,\cdots,i_p}$ is isometric to $O_{1,\cdots,p}$ for any permutation $(i_1, \cdots, i_p) \in \mathcal{S}_p$. This leads to a natural bijection between the set of isometric $(p-1)$-orthoschemes whose union is S_{reg} and the set $\mathcal{S}_p$.

Without loss of generality, let us consider the orthoscheme $O \coloneqq O_{1,\cdots,p}$. Denote by $F_{1,\cdots,k}$ the facet of O opposite to $v_{1,\cdots,k}$. Now move the vertex $\boldsymbol{v}_1$ along the edge $[\boldsymbol{v}_1, \boldsymbol{v}_{1,\cdots,p}]$ towards the vertex $\boldsymbol{v}_{1,\cdots,p}$, to a point $\boldsymbol{s} \in [\boldsymbol{v}_1, \boldsymbol{v}_{1,\cdots,p}]$. Let T be the simplex given by

$$T \coloneqq \operatorname{conv}(\boldsymbol{s}, \boldsymbol{v}_{1,2}, \cdots, \boldsymbol{v}_{1,\cdots,n+1})\,. \tag{3.31}$$

For a given proportionality coefficient $\kappa \in [0,1]$, we want to know where to set $\boldsymbol{s} = \boldsymbol{s}(\kappa)$ on $[\boldsymbol{v}_1, \boldsymbol{v}_{1,\cdots,p}]$ so that

$$\operatorname{vol}_{p-1}(T) = \kappa \cdot \operatorname{vol}_{p-1}(O)\,. \tag{3.32}$$

It is clear that $\kappa = 0 \Leftrightarrow \boldsymbol{s} = \boldsymbol{v}_{1,\cdots,p}$ and that $\kappa = 1 \Leftrightarrow \boldsymbol{s} = \boldsymbol{v}_1$.

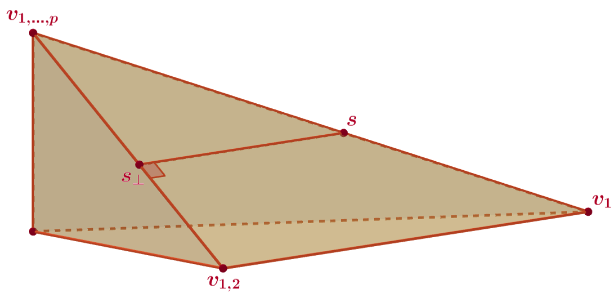
Denote by $\boldsymbol{s}_\perp$ the orthogonal projection of $\boldsymbol{s}$ on F_1, that is, $\boldsymbol{s}_\perp \in F_1$ and $[\boldsymbol{s}, \boldsymbol{s}_\perp] \perp F_1$. The edge $[\boldsymbol{v}_1, \boldsymbol{v}_{1,2}]$ of O is orthogonal to F_1, since

$$[\boldsymbol{v}_1, \boldsymbol{v}_{1,2}] = \bigcap_{k=3}^{p} F_{1,2,\cdots,k}\,, \tag{3.33}$$

and since $F_{1,\cdots,k}$ is orthogonal to F_1 for $3 \le k \le p$ (because O is an orthoscheme). See Fig. 3.12 for an illustration of these observations.

Let F_s be the facet of T opposite to $\boldsymbol{s}$. Because the segments $[\boldsymbol{v}_1, \boldsymbol{v}_{1,2}]$ and $[\boldsymbol{s}, \boldsymbol{s}_\perp]$ are orthogonal to F_1 and F_s respectively, the volumes of O and T are given by

$$\operatorname{vol}_{p-1}(O) = \frac{1}{p-1} \cdot \operatorname{dist}(\boldsymbol{v}_1, \boldsymbol{v}_{1,2}) \cdot \operatorname{vol}_{p-2}(F_1)\,, \tag{3.34}$$



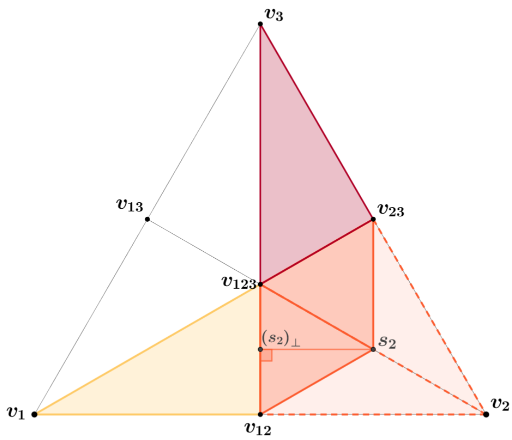


Figure 3.12: Left: Vertex s in the segment $[v_1, v_{1,\cdots,p}]$ and its projection on the facet F_1. Right: Explicit realization of the construction of the left panel for $p = 3$.

and

$$\mathrm{vol}_{p-1}(T) = \frac{1}{p-1} \cdot \mathrm{dist}(s, s_\perp) \cdot \mathrm{vol}_{p-2}(F_s)\,. \tag{3.35}$$

Now observe that

$$F_1 = \mathrm{conv}(v_{1,2}, \cdots, v_{1,\cdots,p}) = F_s\,, \tag{3.36}$$

so that the quotient of the volumes of O and T is given by

$$\frac{\mathrm{vol}_{p-1}(T)}{\mathrm{vol}_{p-1}(O)} = \frac{\mathrm{dist}(s, s_\perp)}{\mathrm{dist}(v_1, v_{1,2})}\,, \tag{3.37}$$

so that Condition (3.32) is equivalent to

$$\mathrm{dist}(s, s_\perp) = \kappa \cdot \mathrm{dist}(v_1, v_{1,2})\,. \tag{3.38}$$

Let us now investigate further the relative positions of the segments $[v_1, v_{1,2}]$ and $[s, s_\perp]$. Since $\dim(F_1) = p-2$ (because F_1 is the convex hull of $p-1$ vertices) and since $[v_1, v_{1,2}] \perp F_1$ (as stated above), it follows that the segments $[s, s_\perp]$ and $[v_1, v_{1,2}]$ are parallel. Moreover, by construction, the points v_1, $v_{1,2}$ and s all belong to the 2-face

$$\triangle := \mathrm{conv}(v_1, v_{1,2}, v_{1,\cdots,p})\,, \tag{3.39}$$

of O. Hence, because $[s, s_\perp] \,//\, [v_1, v_{1,2}]$, it follows that $s_\perp$ is contained in $\triangle$ as well. This in turn implies that the point $s_\perp$ is on the segment $[v_{1,2}, v_{1,\cdots,p}]$, since it is an edge of F_1. Let

$$f := \mathrm{conv}(s, s_\perp, v_{1,\cdots,p})\,, \tag{3.40}$$

denote the 2-face of T with vertices s, $s_\perp$ and $v_{1,\cdots,p}$. From the discussion above we deduce that $f \subset \triangle$, and that both are right-angled triangles (f at $s_\perp$, and $\triangle$ at $v_{1,2}$) sharing the vertex $v_{1,\cdots,p}$.

These observations allow us to deduce that Condition (3.38) is equivalent to

$$\mathrm{dist}(s, v_{1,\cdots,p}) = \kappa \cdot \mathrm{dist}(v_1, v_{1,\cdots,p})\,, \tag{3.41}$$

leading us to the *a priori* somewhat intuitive (or seemingly too nice to be true) fact:

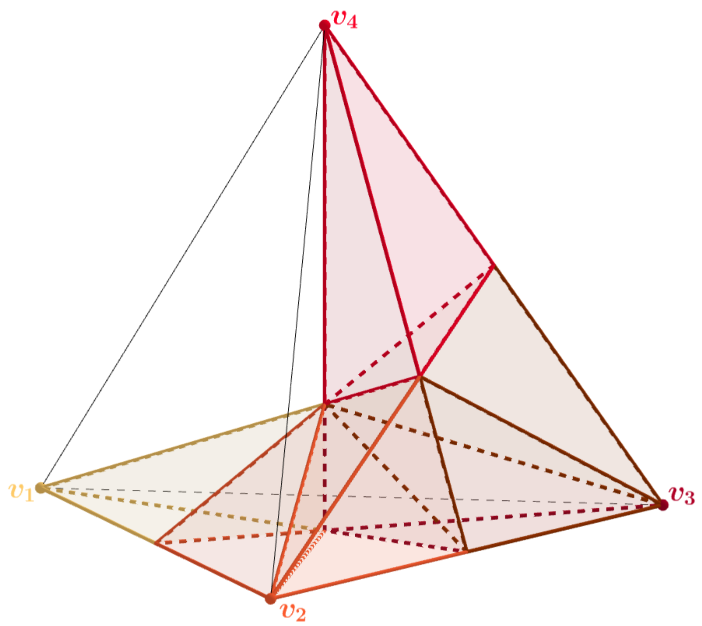


Figure 3.13: Illustration of the set C of admissible opinions for $p = 4$.

Proposition 9. *In order to reduce the volume of O to a proportion $\kappa \in [0,1]$, the vertex $\boldsymbol{v}_1$ has to be moved towards $\boldsymbol{v}_{1,\cdots,p}$ to the point $\boldsymbol{s} \in [\boldsymbol{v}_1, \boldsymbol{v}_{1,\cdots,p}]$ dividing the edge $[\boldsymbol{v}_1, \boldsymbol{v}_{1,\cdots,p}]$ under the same proportion κ.*

Now that we have a way to control the volume of an orthoscheme while keeping one of its facets intact, we turn to the main question of interest for the opinion generation problem.

Without loss of generality, label the parties from 1 to p to match the left-right political spectrum. In order for an opinion to be consistent, we suppose that party i is the most favored (for some $i \in \{1, \cdots, p\}$), and that the adhesion to party j is smaller than (respectively, greater than) the adhesion to party $j+1$ if $j \in \{1, \cdots, i-1\}$ (respectively, $j \in \{i, \cdots, p\}$). These constraints reflect the assumption that as parties are further from the agent's favorite party on the political spectrum, the agent's adhesion towards them decreases. Let $\boldsymbol{x} \in S_{\text{reg}}$ be a point representing the opinion of the agent as follows. The agent favors party j over party k if and only if $\text{dist}(\boldsymbol{x}, \boldsymbol{v}_j) < \text{dist}(\boldsymbol{x}, \boldsymbol{v}_k)$. The set C of all points $\boldsymbol{x} \in S_{\text{reg}}$ representing admissible opinions (in the sense described above) is given by the non-convex polyhedron (see Fig. 3.13)

$$C = \left\{ \boldsymbol{x} \in S_{\text{reg}} \;\middle|\; \begin{array}{c} \text{There is an } i \in \{1, \cdots, n+1\} \text{ such that} \\ \text{dist}(\boldsymbol{x}, \boldsymbol{v}_j) \geq \text{dist}(\boldsymbol{x}, \boldsymbol{v}_{j+1}) \text{ for } j \in \{1, \cdots, i-1\} \text{ and} \\ \text{dist}(\boldsymbol{x}, \boldsymbol{v}_j) \leq \text{dist}(\boldsymbol{x}, \boldsymbol{v}_{j+1}) \text{ for } j \in \{i, \cdots, p-1\} \end{array} \right\}. \tag{3.42}$$

Recall that the regular $(p-1)$-simplex S_{reg} can be decomposed into $p!$ isometric orthoschemes thanks to the barycentric decomposition. Each vertex $\boldsymbol{v}_i$ of S_{reg} is a vertex of $(p-1)!$ such orthoschemes. Let R_i be the union of the $(p-1)!$ orthoschemes having $\boldsymbol{v}_i$ as a vertex. Then $\boldsymbol{x} \in R_i$ if and only if party i is the agent's favorite party. Hence, a point $\boldsymbol{x} \in R_i$ represents a consistent opinion (in the sense described above) if and only if it satisfies the distance conditions from (3.42) for that i. For instance, $C \cap R_1 = O$, the orthoscheme with vertices $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{1,\cdots,p}$ described above.

Recall that the permutations in $\mathcal{S}_p$ are in 1-1 correspondence with the orthoschemes in the barycentric decomposition of S_{reg}, which in turns provides a 1-1 correspondence with all possible ordering of the parties: the permutation $(i_1, \cdots, i_p)$ corresponds to the orthoscheme $\text{conv}(\boldsymbol{v}_{i_1}, \boldsymbol{v}_{i_1,i_2}, \cdots, \boldsymbol{v}_{i_1,\cdots,i_p})$, which corresponds to the opinion with party i_1 being the favorite party, party i_2 the next favorite party (since $\boldsymbol{v}_{i_1,i_2}$ is the midpoint of the edge $[\boldsymbol{v}_{i_1}, \boldsymbol{v}_{i_2}]$), and party i_p the least favorite party.

Conversely, any party ordering corresponds to a permutation in $\mathcal{S}_p$. Therefore, an opinion is consistent (in the sense described above) if and only if its corresponding permutation $(i_1, \cdots, i_p)$ satisfies the following conditions:

$$i_k < i_l < i_1 \Rightarrow k > l\,, \qquad \text{and} \qquad i_k > i_l > i_1 \Rightarrow k > l\,. \tag{3.43}$$

In other words, in the sequence of indices $i_1, \cdots, i_p$, the elements of the subsequences $i_1 - 1, \cdots, 1$ and $i_1 + 1, \cdots, p$ must appear precisely in that order, but elements from different subsequences can be permuted. There are $\binom{p-1}{i_1-1}$ such permutations starting with i_1. Hence, for all $i \in \{1, \cdots, p\}$, $C \cap R_i$ consists in the $\binom{p-1}{i-1}$ orthoschemes corresponding to consistent opinions. In particular, $C \cap R_1$ is the orthoscheme $O_{1,\cdots,p}$ and $C \cap R_p$ is the orthoscheme $O_{p,\cdots,1}$. It follows that the set C of points in S_{reg} corresponding to consistent opinions is a non-convex polyhedron, obtained as the union of

$$\sum_{i=1}^{p} \binom{p-1}{i-1} = 2^{p-1}\,, \tag{3.44}$$

isometric orthoschemes described above. The vertices $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p$ form a subset of the vertices of C, and the vertex $\boldsymbol{v}_i$ is a vertex of $\binom{p-1}{i-1}$ such orthoschemes.

In order for the opinion generation process to be unbiased, the probability to generate an opinion representative $\boldsymbol{x} \in C \cap R_i$ should be the same for all $i \in \{1, \cdots, p\}$. The discussion above provides a natural way to generate such opinions. Observe that for each permutation $(i_1, \cdots, i_p)$, the point $\boldsymbol{v}_{i_1,\cdots,i_p}$ is the center $\boldsymbol{c}$ of S_{reg}. For each vertex $\boldsymbol{v}_i$, move it along the ray $[\boldsymbol{v}_i, \boldsymbol{c}]$ towards $\boldsymbol{c}$ to the point $\boldsymbol{s}_i$ such that

$$\text{dist}(\boldsymbol{s}_i, \boldsymbol{c}) = \frac{1}{\binom{p-1}{i-1}} \cdot \text{dist}(\boldsymbol{v}_i, \boldsymbol{c})\,. \tag{3.45}$$

Notice that this procedure preserves the existing interfaces between $C \cap R_i$ and $C \cap R_j$ for all $i \neq j \in \{1, \cdots, p\}$, so that the non-convex polyhedron C_{r} obtained that way can be seen a non-uniform radial retractation of C with respect of $\boldsymbol{c}$. The set C_{r} is the colored area in the right panel of Fig. 3.12.

Because of Proposition 9, the volumes of all orthoschemes building $C \cap R_i$ are reduced by a factor $\binom{p-1}{i-1}$, so that each of the resulting regions $C_{\text{r}} \cap R_i$ has the same volume, which is equal to the volume of the isometric orthoschemes $O_{1,\cdots,p}$ and $O_{p,\cdots,1}$ (one has indeed that $C \cap R_1 = C_{\text{r}} \cap R_1$ and $C \cap R_p = C_{\text{r}} \cap R_p$).

We summarize the process for generating a random opinion, drawn uniformly in the admissible opinion space:

1. Take the barycentric decomposition of the standard regular simplex S_{reg} into $p!$ isometric orthoschemes.

2. From these orthoschemes, only consider the 2^{p-1} ones corresponding to consistent opinions (their union is the polyhedron C).

3. For all $i \in \{1, \cdots, p\}$, retract the orthoschemes in C containing the vertex $\boldsymbol{v}_i$ (their union is the region R_i) radially with respect to the center $\boldsymbol{c}$ of S_{reg} from a factor $\binom{p-1}{i-1}$. The union of these retracted orthoschemes forms the polyhedron C_{r}. As a result, $\text{vol}_{p-1}(C_{\text{r}} \cap R_i) = \text{vol}_{p-1}(C_{\text{r}} \cap R_j)$ for all $i \neq j \in \{1, \cdots, p\}$.

4. In C_{r}, randomly pick N points, where N is the desired amount of opinions to be generated. The probability that a point $x \in C_{\text{r}}$ belongs to the region $C_{\text{r}} \cap R_i$ is the same for all $i \in \{1, \cdots, p\}$.

4 Conclusion

We attempted to study the robustness of elections against external attacks and identify the relevant factors involved. In particular, we examined how robust elections are depending on (i) social characteristics of the population and (ii) different electoral systems on robustness. We constructed a simple and intuitive opinion dynamics model for studying the evolution of opinions by interaction among individuals. Our model also factors in the impact of an external influence trying to manipulate the opinions. A novelty in our approach is using a real number in an interval to represent an individual's opinion, thereby allowing a certain degree of disagreement in opinion among them. We consider this to be more realistic than constraining individuals to choose one of the few opinions. In addition to capturing the interaction between individuals, our model also encapsulates their natural opinions. We model external influence as a bias to this natural opinion. We let the opinions of individuals evolve in time under our model, to eventually reach an equilibrium state. We call this the final opinion and it is used to decide on whom to vote.

An opinion dynamics model could be said to predict the real trend of human behavior only when we validate it against some actual data. We validated our model by considering how the model can depict the true picture of the volatility of US House of Representatives elections over the years $2012 - 2020$. For this, we introduced a novel concept that we call the *volatility* of elections. We notice a great extent of correlation between the volatility estimated from the real data and from the simulations. This to some extend is a vindication of the validity of our model in studying the robustness of electoral systems.

We presented a detailed validation of our model for two-party system of election. We draw the initial opinion of individuals from a class of multimodal distributions. We tune the parameters of the distribution to capture several properties of the collective opinion of the population such as bias and polarization. We run experiments with different combinations of these parameter values and validate the data obtained against real data by the aforementioned procedure using volatility. The validation experiments showed that the unshifted polarized distribution with biased weight has a higher Pearson correlation coefficient with real data, than shifted polarized distribution with equal weights and shifted polarized distribution with equal weights, concluding that unshifted polarized distribution with biased weight could portray the actual election more effectively than the other distributions. This also gives hints about how the initial opinion is actually distributed among the population.

Using our model, we study how the different characteristics of initial opinion distribution impacts the robustness of election processes. Our model showed that the electoral process is systematically more robust for a two-party system as the communication distance ϵ increases, which drives the connectivity between agents. We also observed that the robustness of an election outcome is favored by less polarization and biasing of the population. Surprisingly, a polarized society is less robust to external influence than a less polarized society. Overall, we observed that a population becomes more resilient to external influence and manipulation by promoting communication and interaction among agents. This is reflected in the increased robustness observed when the confidence bound ϵ is higher. When agents have more opportunities to exchange and share their opinions, it becomes more challenging for external forces to sway the overall outcome of the election.

Another aspect we investigate is the comparison of the robustness of different electoral systems. In a two-party election setup, we consider three different electoral systems, namely Proportional representation (PR), Single representative (SR), and Winner takes all (WTA). For each electoral system, we performed our experiments on 15 synthetically generated countries and corresponding voting population. We observed that PR electoral system consistently outperforms the other two in terms of robustness to external influences. We note that this superiority of PR electoral system is irrespective of the natural distribution employed.

The model is then extended to include the electoral system with more than two parties. Our study mainly focused on the plurality and ranked-choice voting system in a single electoral unit and its variations when considering multiple electoral units. Moreover, the natural opinions we considered have some additional constraints for a multi-partite system that we believe apply to a realistic opinion of an agent. Specifically, we impose that the political affinity of an agent is maximum at some point in the left-right political spectrum and decreases in either sides of the maximum.

For the plurality system, with an almost equal number of agents in each party, all parties have an almost equal percentage of wins for small values of ϵ. But when the interaction between agents increases, the percentage of wins increases for the parties at the center and decreases for the extremist parties because of their corresponding interactions. Contrary to the two-party system, the robustness of the multi-party system has a non-monotonous behavior as a function of confidence bound ϵ. The system's robustness reaches a maximum at ϵ equal to ϵ_m and then decreases. Initially, for small values of ϵ, each party's agents interact within themselves rather than with the agents of other parties. Thus the effort needed to change the election outcome increases as a function of ϵ. However, with a further increase in ϵ, more interconnections among parties help easily exchange ideas, and less effort is needed to change the election outcome. The effort needed to change the election outcome in favor of the extremists is significantly higher compared to the non-extremist parties. Thus the overall effort needed to change the election outcome in favor of the first runner is much lower for larger values of ϵ, where the probability of extremists being the first runner-up is nearly zero.

For the ranked choice voting system, the interactions between the agents and their second-priority votes of the eliminated party determine the percentage of wins. Similar to the plurality system, the RCV system's effort curve is driven by the most extreme party based on the position that the first runner-up belongs on the left-right spectrum. This confirms our analysis of the plurality system.

When comparing different types of electoral systems over multiple electoral units, we find that the PR electoral system is the most reliable, followed by the PRCV, WTAR, and SR electoral systems. We conclude that PR and SR electoral systems are the most and least stable respectively.

Future Work

Our work could be extended in several ways. One of those approaches could be accounting for other systems, for example, two-round systems, block voting, etc. This will allow researchers to understand how various voting mechanisms impact political dynamics and representation. Another possible extension could be to study different influence strategies in the electoral systems.

Delving into different influence strategies within electoral systems opens the door to analyzing how political actors navigate voting to advance their agendas. This can shed light on the effectiveness of different campaigning tactics and the manipulation of voter sentiment, helping policymakers and election monitors develop strategies to maintain fairness and transparency in elections. A study on the effect of heterogeneous values of ϵ allows for exploring scenarios where voters have varying levels of open-mindedness or resistance to change. Understanding how such diversity impacts electoral outcomes can lead to designing voting systems that accommodate different perspectives and promote inclusivity. Another interesting work would be to study how to mitigate the effects of external influences such as media or computational propaganda. By investigating the potential vulnerabilities of electoral systems to external manipulation, researchers can develop strategies to fortify democratic processes, ensuring that citizens are well-informed and free from undue external influences that may distort their voting decisions.

Bibliography

[1] G. M. Givi, R. Delabays, M. Jacquemet, and P. Jacquod, "On the robustness of democratic electoral processes to computational propaganda", Scientific Reports **14** (2024).

[2] J. Nolan, D. Parker, G. C. Van Kooten, and T. Berger, "An overview of computational modeling in agricultural and resource economics", Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie **57**, 417–429 (2009).

[3] A. Bovet and H. A. Makse, "Influence of fake news in twitter during the 2016 US presidential election", Nature communications **10**, 7 (2019).

[4] M. Hoferer, L. Böttcher, H. J. Herrmann, and H. Gersbach, "The impact of technologies in political campaigns", Physica A **538**, 122795 (2020).

[5] S. H. Strogatz, *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering* (Westview Press, 2000).

[6] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics", Rev. Mod. Phys. **81**, 591–646 (2009).

[7] M. Buchanan, *The social atom: why the rich get richer, cheaters get caught, and your neighbour usually looks like you* (Jan. 2007).

[8] A. Sîrbu, V. Loreto, V. D. P. Servedio, and F. Tria, "Opinion dynamics: models, extensions and external effects", in *Participatory sensing, opinions and collective awareness*, edited by V. Loreto, M. Haklay, A. Hotho, V. D. Servedio, G. Stumme, J. Theunis, and F. Tria (Springer International Publishing, Cham, 2017), pp. 363–401.

[9] Y. Dong, M. Zhan, G. Kou, Z. Ding, and H. Liang, "A survey on the fusion process in opinion dynamics", Inf. Fusion **43**, 57–65 (2018).

[10] J. R. P. French, "A formal theory of social power.", Psycho. Rev. **63**, 181–194 (1956).

[11] M. H. DeGroot, "Reaching a consensus", J. Amer. Stat. Assoc. **69**, 118–121 (1974).

[12] R. Abelson, "Mathematical models of the distribution of attitudes under controversy", in *Contributions to mathematical psychology*, edited by N. Fredericksen and H. Gullicksen (Holt, Rinehart & Winston, 1964).

[13] M. Taylor, "Towards a mathematical theory of influence and attitude change", Human Rel. **21**, 121–139 (1968).

[14] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions", J. Math. Socio. **15**, 193–206 (1990).

[15] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation", J. Artif. Soc. Social Simul. **5**, 1–2 (2002).

[16] J. Lorenz, "Continuous Opinion Dynamics under Bounded Confidence: A Survey", Int. J. Modern Phys. C **18**, 1819–1838 (2007).

[17] T. M. Liggett and T. M. Liggett, *Interacting particle systems*, Vol. 2 (Springer, 1985).

[18] P. N. Howard, S. Woolley, and R. Calo, “Algorithms, bots, and political communication in the us 2016 election: the challenge of automated political communication for election law and administration”, J. Inf. Technol. Politics **15**, 81–93 (2018).

[19] E. Landgren, J. L. Juul, and S. H. Strogatz, “How a minority can win: unrepresentative outcomes in a simple model of voter turnout”, Physical Review E **104**, 10.1103/physreve.104.054307 (2021).

[20] K. Breuninger, *Russia tried to influence u.s. elections in 2022 and will do it again, nation’s top intel agency says.*

[21] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, “Experimental evidence of massive-scale emotional contagion through social networks”, Proceedings of the National Academy of Sciences **111**, 8788 (2014).

[22] P. T. Metaxas and E. Mustafaraj, “Social media and the elections”, Science **338**, 472 (2012).

[23] M. Cai, H. Luo, X. Meng, Y. Cui, and W. Wang, “Network distribution and sentiment interaction: information diffusion mechanisms between social bots and human users on social media”, Information Processing and Management **60**, 103197 (2023).

[24] F. Harary, “A criterion for unanimity in french’s theory of social power”, In D. Cartwright (Ed.), Studies in social power, 168–182 (1959).

[25] R. L. Berger, “A necessary and sufficient condition for reaching a consensus using degroot’s method”, Journal of the American Statistical Association **76**, 415–418 (1981).

[26] N. Friedkin and E. Johnsen, “Social influence networks and opinion change”, Advances in Group Processes **16** (1999).

[27] S. Parsegov, A. Proskurnikov, R. Tempo, and N. Friedkin, “Novel multidimensional models of opinion dynamics in social networks”, IEEE Transactions on Automatic Control **62**, 2270– (2017).

[28] F. Baumann, I. M. Sokolov, and M. Tyloo, “A Laplacian approach to stubborn agents and their role in opinion formation on influence networks”, Physica A **557**, 124869 (2020).

[29] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, “Mixing beliefs among interacting agents”, Adv. Complex Syst. **03**, 87–98 (2000).

[30] G. Kou, Y. Zhao, Y. Peng, and Y. Shi, “Multi-level opinion dynamics under bounded confidence”, PLOS ONE **7**, 1–10 (2012).

[31] S. Fortunato, “Universality of the threshold for complete consensus for the opinion dynamics of Deffuant et al.”, International Journal of Modern Physics C **15**, 1301–1307 (2004).

[32] G. Deffuant, F. Amblard, and G. Weisbuch, “How can extremism prevail? A study based on the relative agreement interaction model”, Journal of Artificial Societies and Social Simulation **5**, 1–1 (2002).

[33] F. Amblard and G. Deffuant, “The role of network topology on extremism propagation with the relative agreement opinion dynamics”, Physica A: Statistical Mechanics and its Applications **343**, 725–738 (2004).

[34] D. Stauffer, A. Sousa, and C. Schulz, "Discretized opinion dynamics of the Deffuant model on scale-free networks", Journal of Artificial Societies and Social Simulation **7** (2004).

[35] J. Lorenz, "Heterogeneous bounds of confidence: meet, discuss and find consensus!", Complexity, NA–NA (2009).

[36] G. Weisbuch, G. Deffuant, F. Amblard, and J.-P. Nadal, "Meet, discuss, and segregate!", Complexity **7**, 55–63 (2002).

[37] J. Lorenz, "Repeated averaging and bounded confidence modeling, analysis and simulation of continuous opinion dynamics", PhD thesis (2007).

[38] E. Ben-Naim, P. Krapivsky, and S. Redner, "Bifurcations and patterns in compromise processes", Physica D: Nonlinear Phenomena **183**, 190–204 (2003).

[39] U. Krause, "A discrete nonlinear and non-autonomous model of consensus formation", Communications in Difference Equations **2000**, 10.1201/b16999-21 (2000).

[40] A. Bhattacharyya, M. Braverman, B. Chazelle, and H. L. Nguyen, "On the convergence of the Hegselmann-Krause system", CoRR **abs/1211.1909** (2012).

[41] F. Slanina, "Dynamical phase transitions in Hegselmann-Krause model of opinion dynamics and consensus", The European Physical Journal B **79**, 99–106 (2010).

[42] V. D. Blondel, J. M. Hendrickx, and J. N. Tsitsiklis, "On Krause's multi-agent consensus model with state-dependent connectivity", IEEE transactions on Automatic Control **54**, 2586–2597 (2009).

[43] S. Fortunato, "On the consensus threshold for the opinion dynamics of Krause-Hegselmann", Int. J. Modern Phys. C **16**, 259–270 (2004).

[44] J. Lorenz, *Continuous opinion dynamics: insights through interactive markov chains*, 2007.

[45] A. Pluchino, V. Latora, and A. Rapisarda, "Compromise and synchronization in opinion dynamics", The European Physical Journal B - Condensed Matter and Complex Systems **50**, 169–176 (2006).

[46] S. Fortunato, V. Latora, A. Pluchino, and A. Rapisarda, "Vector opinion dynamics in a bounded confidence consensus model", International Journal of Modern Physics C **16**, 1535–1551 (2005).

[47] A. Nedić and B. Touri, "Multi-dimensional Hegselmann-Krause dynamics", in 2012 IEEE 51st IEEE conference on decision and control (CDC) (2012), pp. 68–73.

[48] J. Lorenz, "Multidimensional opinion dynamics when confidence changes", (2003).

[49] P. Clifford and A. Sudbury, "A model for spatial conflict", Biometrika **60**, 581–588 (1973).

[50] R. A. Holley and T. M. Liggett, "Ergodic theorems for weakly interacting infinite systems and the voter model", The Annals of Probability **3**, 643–663 (1975).

[51] J. T. Cox, "Coalescing random walks and voter model consensus times on the torus in zd", The Annals of Probability, 1333–1366 (1989).

[52] D. Volovik and S. Redner, "Dynamics of confident voting", J. Stat. Mech. **2012**, P04003 (2012).

[53] M. Mobilia, “Does a single zealot affect an infinite group of voters?”, Phys. Rev. Lett. **91**, 028701 (2003).

[54] M. Mobilia, A. Petersen, and S. Redner, “On the role of zealotry in the voter model”, Journal of Statistical Mechanics: Theory and Experiment **2007**, P08029 (2007).

[55] F. Vazquez and S. Redner, “Ultimate fate of constrained voters”, Journal of Physics A: Mathematical and General **37**, 8479–8494 (2004).

[56] M. Mobilia, “Fixation and polarization in a three-species opinion dynamics model”, Europhysics Letters **95**, 50002 (2011).

[57] M. Mobilia, “Commitment versus persuasion in the three-party constrained voter model”, Journal of Statistical Physics **151**, 69–91 (2013).

[58] C. Castellano, M. A. Muñ oz, and R. Pastor-Satorras, “The non-linear q-voter model”, Physical Review E **80**, 10.1103/physreve.80.041129 (2009).

[59] P. Moretti, S. Liu, C. Castellano, and R. Pastor-Satorras, “Mean-field analysis of the q-voter model on networks”, Journal of statistical physics **151**, 113–130 (2013).

[60] P. Nyczka, K. Sznajd-Weron, and J. Cisło, “Phase transitions in the q-voter model with two types of stochastic driving”, Phys. Rev. E **86**, 011105 (2012).

[61] S. Galam, “Minority opinion spreading in random geometry”, Physics of Condensed Matter **25**, 10.1140/epjb/e20020045 (2002).

[62] C. J. Tessone, R. Toral, P. Amengual, H. S. Wio, and M. S. Miguel, “Neighborhood models of minority opinion spreading”, The European Physical Journal B **39**, 535–544 (2004).

[63] S. Galam, *Sociophysics and the forming of public opinion: threshold versus non threshold dynamics*, 2008.

[64] S. Galam, “Heterogeneous beliefs, segregation, and extremism in the making of public opinions”, Phys. Rev. E **71**, 046123 (2005).

[65] S. Galam and F. Jacobs, “The role of inflexible minorities in the breaking of democratic opinion dynamics”, Physica A: Statistical Mechanics and its Applications **381**, 366–376 (2007).

[66] S. Galam, “Public debates driven by incomplete scientific data: the cases of evolution theory, global warming and h1n1 pandemic influenza”, ArXiv **abs/1004.5009** (2010).

[67] Z.-X. Wu and G. Chen, *Effects of free will and massive opinion in majority rule model*, 2008.

[68] S. Galam, “Collective beliefs and individual stubbornness in the dynamics of public debates”, arXiv preprint arXiv:1006.2332 (2010).

[69] S. Galam, *Market efficiency, anticipation and the formation of bubbles-crashes*, 2011.

[70] N. Lanchier and J. Neufer, “Stochastic dynamics on hypergraphs and the spatial majority rule model”, Journal of Statistical Physics **151**, 21–45 (2013).

[71] Q. Li, L. A. Braunstein, H. Wang, J. Shao, H. E. Stanley, and S. Havlin, “Non-consensus opinion models on complex networks”, Journal of Statistical Physics **151**, 92–112 (2013).

[72] S. Galam, “Sociophysics: A review of Galam models”, Int. J. Modern Phys. C **19**, 409–440 (2008).

[73] K. Sznajd-Weron and J. Sznajd, "Opinion evolution in closed community", Int. J. Modern Phys. C **11**, 1157–1165 (2000).

[74] K. Sznajd-Weron, "Sznajd model and its applications", Acta Physica Polonica B **36** (2005).

[75] K. Sznajd-Weron, M. Tabiszewski, and A. M. Timpanaro, "Phase transition in the Sznajd model with independence", Europhysics Letters **96**, 48002 (2011).

[76] C. Schulze, "Advertising in the Sznajd marketing model", International Journal of Modern Physics C **14**, 95–98 (2003).

[77] K. Sznajd-Weron and R. Weron, "A simple model of price formation", International Journal of Modern Physics C **13**, 115–123 (2002).

[78] R. N. C. Filho, M. P. Almeida, J. S. Andrade, and J. E. Moreira, "Scaling behavior in a proportional voting process", Phys. Rev. E **60**, 1067–1068 (1999).

[79] L. Behera and F. Schweitzer, "On spatial consensus formation: is the Sznajd model different from a voter model?", International Journal of Modern Physics C **14**, 1331–1354 (2003).

[80] D. Stauffer, "Difficulty for consensus in simultaneous opinion formation of Sznajd model", Mathematical Sociology **28**, 25–33 (2004).

[81] D. Stauffer, A. Sousa, and S. Oliveira, "Generalization to square lattice of Sznajd sociophysics model", International Journal of Modern Physics C - IJMPC **11**, 1239–1245 (2000).

[82] A. Bernardes, D. Stauffer, and J. Kertész, "Election results and the Sznajd model on barabasi network", The European Physical Journal B - Condensed Matter **25**, 123–127 (2002).

[83] F. Slanina and H. Lavicka, "Analytical results for the Sznajd model of opinion formation", The European Physical Journal B-Condensed Matter and Complex Systems **35**, 279–288 (2003).

[84] G. Kondrat and K. Sznajd-Weron, "Spontaneous reorientations in a model of opinion dynamics with anticonformists", International Journal of Modern Physics C **21**, 559–566 (2010).

[85] M. Rybak and K. Kulakowski, "Competing of Sznajd and voter dynamics in the watts-strogatz network", arXiv preprint arXiv:1301.2944 (2013).

[86] R. Axelrod, "The dissemination of culture: a model with local convergence and global polarization", Journal of conflict resolution **41**, 203–226 (1997).

[87] C. Castellano, M. Marsili, and A. Vespignani, "Nonequilibrium phase transition in a model for social influence", Phys. Rev. Lett. **85**, 3536–3539 (2000).

[88] K. Klemm, V. M. Eguíluz, R. Toral, and M. S. Miguel, "Global culture: a noise-induced transition in finite systems", Phys. Rev. E **67**, 045101 (2003).

[89] J. C. Gonzélez-Avella, M. G. Cosenza, and K. Tucci, "Nonequilibrium transition induced by mass media in a model for social influence", Phys. Rev. E **72**, 065102 (2005).

[90] N. Lanchier, "The Axelrod model for the dissemination of culture revisited", The Annals of Applied Probability **22**, 860–880 (2012).

[91] N. Lanchier and S. Scarlatos, "Fixation in the one-dimensional Axelrod model", The Annals of Applied Probability **23**, 2538–2559 (2013).

[92] D. Kempe, J. Kleinberg, S. Oren, and A. Slivkins, *Selection and influence in cultural dynamics*, 2013.

[93] L. A. Barbosa and J. F. Fontanari, "Culture–area relation in Axelrod's model for culture dissemination", Theory in Biosciences **128**, 205–210 (2009).

[94] M. Laguna, G. Abramson, and D. H. Zanette, "Vector opinion dynamics in a model for social influence", Physica A: Statistical Mechanics and its Applications **329**, 459–472 (2003).

[95] D. Jacobmeier, "Multidimensional consensus model on a Barabàsi-Albert network", International Journal of Modern Physics C **16**, 633–646 (2005).

[96] S. Huet, G. Deffaunt, and W. Jager, "A rejection mechanism in a 2d bounded confidence provides more conformity", Advances in Complex Systems **11**, 529–549 (2008).

[97] J. Lorenz, "Continuous Opinion Dynamics of Multidimensional Allocation Problems under Bounded Confidence. More dimensions lead to better chances for consensus", European Journal of Economic and Social Systems **19**, 213–227 (2006).

[98] D. Helbing, *Managing complexity: insights, concepts, applications* (Jan. 2008).

[99] S. R. Etesami and T. Basar, "Game-theoretic analysis of the Hegselmann-Krause model for opinion dynamics in finite dimensions", CoRR **abs/1412.6546** (2014).

[100] A. Waagen, G. Verma, K. Chan, A. Swami, and R. D'Souza, "Effect of zealotry in high-dimensional opinion dynamics models", Phys. Rev. E **91**, 022811 (2015).

[101] S. T. Stamoulas and M. Rathinam, "Convergence, stability, and robustness of multidimensional opinion dynamics in continuous time", SIAM J. Control Optim. **56**, 1938–1967 (2018).

[102] G. Soda and A. Zaheer, "A network perspective on organizational architecture: performance effects of the interplay of formal and informal organization", Strategic Management Journal **33**, 751–771 (2012).

[103] P. Erdős, A. Rényi, et al., "On the evolution of random graphs", Publ. Math. Inst. Hung. Acad. Sci **5**, 17–60 (1960).

[104] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks", Nature **393**, 440–442 (1998).

[105] A.-L. Barabési and R. Albert, "Emergence of scaling in random networks", Science **286**, 509–512 (1999).

[106] D. Stauffer and H. Meyer-Ortmanns, "Simulation of consensus model of Deffuant et al on a Barabàsi-Albert network", International Journal of Modern Physics C **15**, 241–246 (2004).

[107] Y. Gandica, M. del Castillo-Mussot, G. J. Vázquez, and S. Rojas, "Continuous opinion model in small-world directed networks", Physica A: Statistical Mechanics and its Applications **389**, 5864–5870 (2010).

[108] D. Stauffer and H. Meyer-Ortmanns, "Simulation of consensus model of Deffuant et al. on a Barabàsi-Albert network", International Journal of Modern Physics C **15**, 241–246 (2004).

[109] M. Galesic and D. Stein, “Statistical physics models of belief dynamics: theory and empirical tests”, Physica A: Statistical Mechanics and its Applications **519**, 275–294 (2019).

[110] A. F. Peralta, J. Kertész, and G. Iñiguez, “Opinion dynamics in social networks: from models to data”, arXiv preprint arXiv:2201.01322 (2022).

[111] V. Pansanella, V. Morini, T. Squartini, and G. Rossetti, *Change my mind: data driven estimate of open-mindedness from political discussions*, 2022.

[112] H. C. Kelman, “Compliance, identification, and internalization three processes of attitude change”, J. Conflict Res. **2**, 51–60 (1958).

[113] S. Balietti, L. Getoor, D. G. Goldstein, and D. J. Watts, “Reducing opinion polarization: effects of exposure to similar people with differing political views”, Proceedings of the National Academy of Sciences **118**, e2112552118 (2021).

[114] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election”, Journal of economic perspectives **31**, 211–36 (2017).

[115] A. T. Bernardes, U. M. S. Costa, A. D. Araujo, and D. Stauffer, “Damage spreading, coarsening dynamics and distribution of polititcal votes in Sznajd model on square lattice”, International Journal of Modern Physics C **12**, 159–167 (2001).

[116] M. G. lez, A. Sousa, and H. Herrmann, “Opinion formation on a deterministic pseudo-fractal network”, International Journal of Modern Physics C **15**, 45–57 (2004).

[117] S. Fortunato and C. Castellano, “Scaling and universality in proportional elections”, Physical Review Letters **99**, 10.1103/physrevlett.99.138701 (2007).

[118] T. Yasseri and J. Bright, “Wikipedia traffic data and electoral prediction: towards theoretically informed models”, EPJ Data Science **5**, 10.1140/epjds/s13688-016-0083-3 (2016).

[119] J. Fernández-Gracia, K. Suchecki, J. J. Ramasco, M. S. Miguel, and V. M. Eguíluz, “Is the voter model a model for voters?”, Physical Review Letters **112**, 10.1103/physrevlett.112.158701 (2014).

[120] A. Kononovicius, “Empirical analysis and agent-based modeling of the lithuanian parliamentary elections”, Complexity **2017**, 1–15 (2017).

[121] A. Kononovicius, “Modeling of the parties’ vote share distributions”, Acta Physica Polonica A **133**, 1450–1458 (2018).

[122] T. Phan, M. Navas-Zuloage, C. Ignace, C. Castillo-Garsow, V. Moreno, and C. Kribs, “Modeling voting dynamics in a two-party system”, Revista de Matemática: Teoría y Aplicaciones **27**, 179–219 (2019).

[123] N. Crokidakis and P. M. C. de Oliveira, “The Sznajd model with limited persuasion: competition between high-reputation and hesitant agents”, Journal of Statistical Mechanics: Theory and Experiment **2011**, P11004 (2011).

[124] K. Sznajd-Weron, R. Weron, and M. Włoszczowska, “Outflow dynamics in modeling oligopoly markets: the case of the mobile telecommunications market in poland”, Journal of Statistical Mechanics: Theory and Experiment **2008**, P11018 (2008).

[125] J. C. González-Avella, V. M. Eguíluz, M. Marsili, F. Vega-Redondo, and M. San Miguel, “Threshold learning dynamics in social networks”, PLOS ONE **6**, 1–9 (2011).

[126] M. Fabiana Laguna, G. Abramson, and J. R. Iglesias, “Compelled to do the right thing”, The European Physical Journal B **86**, 1–8 (2013).

[127] T. Carletti, D. Fanelli, S. Grolli, and A. Guarino, “How to make an efficient propaganda”, Europhys. Lett. **74**, 222–228 (2006).

[128] F. Gargiulo, S. Lottini, and A. Mazzoni, “The saturation threshold of public opinion: are aggressive media campaigns always effective?”, 10.48550/ARXIV.0807.3937 (2008).

[129] T. Vaz Martins, M. Pineda, and R. Toral, “Mass media and repulsive interactions in continuous-opinion dynamics”, EPL (Europhysics Letters) **91**, 48003 (2010).

[130] W. Quattrociocchi, G. Caldarelli, and A. Scala, “Opinion dynamics on interacting networks: media competition and social influence”, Scientific reports **4**, 10.1038/srep04938 (2014).

[131] W. Quattrociocchi, R. Conte, and E. Lodi, “Opinions manipulation: media, power and gossip”, Advances in Complex Systems **14**, 567–586 (2011).

[132] R. Hegselmann and U. Krause, “Truth and cognitive division of labour: first steps towards a computer aided social epistemology”, Journal of Artificial Societies and Social Simulation **9** (2006).

[133] S. Kurz and J. Rambau, “On the hegselmann–krause conjecture in opinion dynamics”, Journal of Difference Equations and Applications **17**, 859–876 (2011).

[134] J. Gonzalez-Avella, V. M. Eguiluz, and M. S. Miguel, *Group formation and Mass Media effects in Cultural Dynamics: The power of being subtle*, Computing in Economics and Finance 2006 450 (Society for Computational Economics, July 2006).

[135] J. C. González-Avella, M. G. Cosenza, V. M. Eguíluz, and M. S. Miguel, “Spontaneous ordering against an external field in non-equilibrium systems”, New Journal of Physics **12**, 013010 (2010).

[136] L. R. Peres and J. F. Fontanari, “The mass media destabilizes the cultural homogenous regime in Axelrod’s model”, Journal of Physics A: Mathematical and Theoretical **43**, 055003 (2010).

[137] L. R. Peres and J. F. Fontanari, “The media effect in Axelrod’s model explained”, Europhysics Letters **96**, 38004 (2011).

[138] J. C. González-Avella, M. G. Cosenza, and M. San Miguel, “A model for cross-cultural reciprocal interactions through mass media”, PLOS ONE **7**, 1–7 (2012).

[139] J. C. Gonzélez-Avella, V. M. Eguíluz, M. G. Cosenza, K. Klemm, J. L. Herrera, and M. San Miguel, “Local versus global interactions in nonequilibrium transitions: a model of social dynamics”, Phys. Rev. E **73**, 046119 (2006).

[140] A. H. Rodríguez, M. del Castillo-Mussot, and G. J. Vézquez, “Induced cultural globalization by an external vector field in an enhanced Axelrod model”, in 7th international conference on practical applications of agents and multi-agent systems (paams 2009), edited by Y. Demazeau, J. Pavón, J. M. Corchado, and J. Bajo (2009), pp. 310–318.

[141] K. I. Mazzitello, J. Candia, and V. Dossetti, “Effects of mass media and cultural drift in a model for social influence”, International Journal of Modern Physics C **18**, 1475–1482 (2007).

[142] A. H. Rodríguez and Y. Moreno, "Effects of mass media action on the Axelrod model with social influence", Physical Review E **82**, 016111 (2010).

[143] J. Candia and K. I. Mazzitello, "Mass media influence spreading in social networks with community structure", Journal of Statistical Mechanics: Theory and Experiment **2008**, P07007 (2008).

[144] A. Sîrbu, V. Loreto, V. D. Servedio, and F. Tria, "Opinion dynamics with disagreement and modulated information", Journal of Statistical Physics **151**, 218–237 (2013).

[145] A. Sîrbu, V. Loreto, V. D. Servedio, and F. Tria, "Cohesion, consensus and extreme information in opinion dynamics", Advances in Complex Systems **16**, 1350035 (2013).

[146] *Attacks from within seen as a growing threat to elections.*

[147] S. Z. Akbar, A. Panda, and J. Pal, "Political hazard: misinformation in the 2019 indian general election campaign", South Asian History and Culture **13**, 399–417 (2022).

[148] S. Sanovich, *Computational propaganda in russia: the origins of digital misinformation*, tech. rep. (2017), pp. 1–25.

[149] S. Woolley and D. Guilbeault, *Computational propaganda in the united states of america: manufacturing consensus online*, tech. rep. (2017), pp. 1–29.

[150] P. N. Howard, "The rise of computational propaganda", IEEE Spectrum (2018).

[151] P. N. Howard and B. Kollanyi, "Bots, #strongerin, and #brexit: computational propaganda during the UK-EU referendum", SSRN Electronic Journal (2016).

[152] A. Bessi and E. Ferrara, "Social bots distort the 2016 US presidential election", SSRN Electronic Journal **21** (2016).

[153] D. Arnaudo, "Computational propaganda in brazil: social bots during elections", Computational Propaganda Research Project **8**, 1–39 (2017).

[154] C. de Secondat Baron de Montesquieu, *The spirit of laws* (1748).

[155] FairVote, *Different types of electoral system worldwide*, 2022.

[156] Aceproject, *The electoral knowledge network*, 2023.

[157] G. Chen, W. Su, W. Mei, and F. Bullo, "Convergence properties of the heterogeneous Deffuant–Weisbuch model", Automatica **114**, 108825 (2020).

[158] V. C. Yang, D. M. Abrams, G. Kernell, and A. E. Motter, "Why are u.s. parties so polarized? a "satisficing" dynamical model", SIAM Review **62**, 646–657 (2020).

[159] R. Axelrod, J. J. Daymude, and S. Forrest, "Preventing extreme polarization of political attitudes", Proceedings of the National Academy of Sciences **118**, e2102139118 (2021).

[160] R. A. Horn and C. R. Johnson, *Matrix analysis* (Cambridge University Press, New York, 1994).

[161] R. A. Horn and C. R. Johnson, *Matrix analysis* (Cambridge University Press, New York, 1986).

[162] A. Ostrowski, "über die Determinanten mit überwiegender Hauptdiagonal", Comm. Math. Helv. **10**, 69–96 (1937).

[163] R. J. Plemmons, "M-matrix characterizations I – nonsingular M-matrices", Lin. Alg. Appl. **18**, 175–188 (1977).

[164] H. S. M. Coxeter, *Regular polytopes*, 3d ed (Dover Publications, New York, 1973).