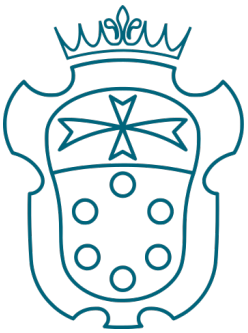


SCUOLA
NORMALE
SUPERIORE

Classe di Scienze
PhD in Data Science

PHD THESIS

# "BIASED ECHOES: UNRAVELING MECHANISMS OF OPINION DYNAMICS AND THEIR IMPACTS IN ONLINE SOCIAL NETWORKS"

SUPERVISORS
Prof. Giulio ROSSETTI
Prof. Tiziano SQUARTINI

CANDIDATE
Valentina PANSANELLA

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy (Ph.D.) in Data Science (XXXVI cycle)

# Declaration of Authorship

I, Valentina Pansanella, declare that this thesis titled, "Biased Echoes: Unraveling Mechanisms of Opinion Dynamics and their Impacts in Online Social Networks" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

January 23, 2024

# Abstract

The societal role played by public and individual opinions is crucial, not only because they shape our culture but also because they drive individual and - indirectly - collective actions, by influencing political decisions. Therefore, it is important to address and solve the problem of understanding how opinions form and evolve (in the context of online social networks), as this has significant implications for opinion dynamics, polarization, and social AI. Despite skepticism and contrasting empirical insights, it is evident that the digital era lead to additional complexities into this process and posed a threat for an healthy process of opinion evolution, contributing to the creation and maintenance of *polluted information environments*. In this thesis, therefore, our aim was to investigate the interplay of biases and network effects in driving opinion formation and diffusion in online social networks. We first review the state-of-the-art in computational social sciences, focusing on the structures used to model societies, such as graphs, temporal graphs, and higher-order structures. We then delve into the milestones of opinion dynamics, discussing models that account for the impact of different underlying structures and characteristic elements of the digital space, such as algorithmic bias. The literature on opinion dynamics is wide, ranging from binary opinions and pair-wise interactions models to continuous opinions on time-evolving higher-order systems, in a never-ending effort to reduce the gap between reality and the models' predictions. However, despite such a rich set of mathematical studies, the works concerning the validation of models on real data are scarce. Our approach was therefore two-fold. First, we developed models of opinion dynamics that incorporate specific characteristics of the opinion formation process and simulate their long-term consequences, i.e. until the studied population reaches an equilibrium, if possible. This thesis places a strong emphasis on capturing the realistic dynamics of online environments by examining the interplay between algorithmic and cognitive biases, which are inherent in all the models under study. These biases are then carefully scrutinized in conjunction with other factors, such as network effects, dynamic influences, and the presence of external agents, including mass media. The resulting models are designed to facilitate the analysis of various scenarios akin to those observed in social media. Additionally, we developed a hybrid approach that leverages existing opinion dynamics models for a time-aware user-level estimate of "open-mindedness" from real online discussions data, using both Reddit and Twitter as a case study. Lastly, we employed such methodology to "validate" one of the proposed model on a real online discussion from Twitter around the Black Lives Matter controversy during Euro2020, introducing a possible pipeline for employing models to explain the unfolding of polluting phenomena on social media. Throughout the work our main focus is to use the simplicity and interpretability of opinion dynamics model to better understand such a complex real phenomena.

# Contents

# List of Figures

# List of Tables

# Chapter 1

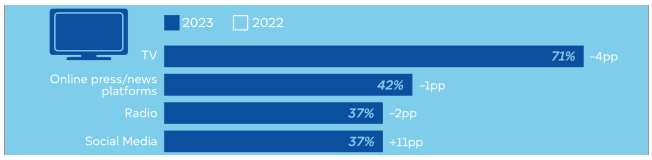
# Introduction

Over the past decade, the pervasiveness of online social media and social networks has rapidly changed how we are accustomed to searching, gathering, and discussing information. However, the unlimited freedom to create content and the unprecedented information overload we are used to today can make online platforms a fertile ground for biased, polluted phenomena. Concerns around **filter bubbles, echo chambers** and (political) **polarization** have remained consistently present in discourses among socio-political scientists for the last twenty years.

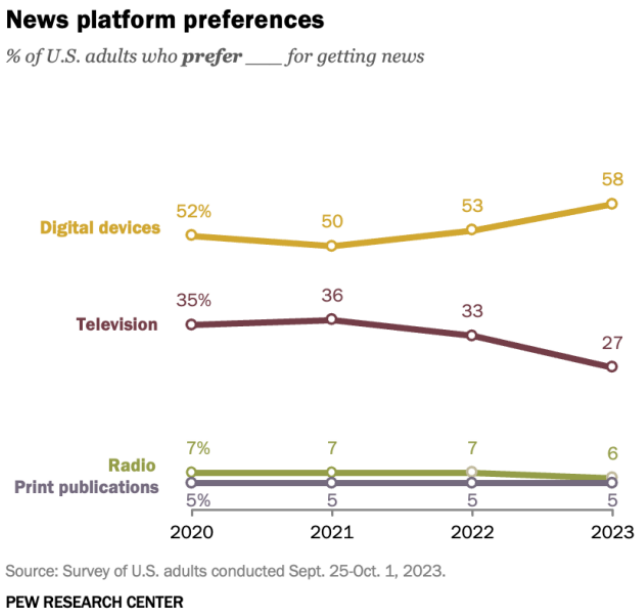
Such major concerns are inevitably interrelated to the dynamics of public opinion (think, for example, to the growth of polarization). This made evident that understanding the process underlying the formation and evolution of opinions is a crucial task that needs to be addressed and solved. Preoccupations are - in fact - that said phenomena may prevent the dialectical process of "thesis-antithesis-synthesis", which is the basis of constructive belief and knowledge formation, with consequences not only limited to the online world.

**What are opinions?** According to the Cambridge American English Dictionary, an opinion is an *idea that a person or a group of people have about something or someone based mainly on their feelings and beliefs*. As this definition points out, an opinion has a highly subjective nature. It is inevitably entangled with psychological, cultural, and social effects on the individual, making the process of public opinion evolution complex and difficult to untangle. To clarify the intricate factors that influence this process, it can be argued that internal and external factors shape an individual's opinion. External factors are rooted in the social and informational environment. It is not uncommon for individuals to be influenced by the opinions of their social network and the information they receive from expert or authoritative sources. On the other hand, internal factors include personal attributes such as cultural background, cognitive biases, and prior beliefs. Humans are - in fact - far from being perfectly rational individuals, and the assumption that opinions form through truth-seeking and rational reasoning is unfortunately not true in most cases.

**Cognitive biases.** With a limited amount of time and attention[108], people choose what to focus on in every area of their lives. Rather than striving for a comprehensive and balanced perspective, people tend to gravitate towards information that aligns with their preexisting beliefs, thereby avoiding the *cognitive dissonance* that arises from encountering contradictory viewpoints. This phenomenon can be attributed to two well-documented cognitive biases: confirmation bias and selection bias. These cognitive biases - alone - are argued to be able to reinforce existing opinions and hinder the development of a more nuanced understanding of complex issues, thus leading to increasing polarization and other polluting phenomena.




(A) Flash Eurobarometer FL012EP – Media & News Survey (2023). *General Media Use in the European Union*




(B) Pew Research Center Survey of US Adults 25 Sept. - 1 Oct. 2023. *News Platform Preferences*

FIGURE 1.1: Statistics on Media consumption in Europe (A) and USA (B) showing the growth of Online Social Media as news sources

**The role of technology.** Online social media and networks are increasingly used as an alternative news source to mainstream media but also as a great arena to express personal opinions, engage in discussion, and share content from other sources with one's network.

According to the 2023 Eurobarometer Media & News Survey report [1], television remains the most common media source among Europeans, but Internet and social media use is steadily rising alongside the decline of newspaper readership (see Chapter 1(A)). Nearly seven out of ten Europeans use online social networks at least once a week, and according to a Pew Research Center survey conducted in 2023, more than half of US adults in 2023 preferred digital services for getting news (see Chapter 1(B)). In the early days, it was argued that the advent of the Internet, guaranteeing free access to a huge amount of information, would be a boon for democracy. Instead, **information proliferation** [106], i.e., the capacity to access and contribute to a growing quantity of information, is reducing the quantity and quality of content many people engage with. In these environments, belief-consistent selection affects the information users choose to interact with and the composition of their network of interactions. While in some platforms, this is mainly determined by real-life social ties (family, friends, colleagues...), in some other contexts, people tend to create their

[1] Flash Eurobarometer FL012EP – Media & News Survey (2023). *General Media Use in the European Union*.

bubble of like-minded individuals - whom they may not even know in real life - to create a comfort zone where there is no disagreement nor conflict.

However, when we talk about online environments, the main concerns about the causes of information environment pollution are not related to users' cognitive behaviors and biases, at least not in the first instance. A major topic of discussion that has opened up in the past decade is the effects that the pervasiveness of algorithms that manage the online experience may have on societal dynamics. The social dimension of Artificial Intelligence (AI) is increasingly present in our daily lives due to the ubiquity of complex socio-technical systems that involve people, algorithms, and machines interacting with each other. Artificial intelligence has the potential to empower individuals in tackling complex societal issues, yet it can also worsen societal problems and vulnerabilities, including bias, inequality, and polarization. For example, the user-driven biased selection process described above is arguably reinforced by the presence of Recommender Systems (RS) and algorithmic filtering, displaying content similar to user-created content and suggesting new connections based on profile similarities. This is argued to create a positive feedback loop, further reducing the amount of diversity in the user experience[2]. - hence possibly contributing to creating and maintaining echo chambers and filter bubbles, phenomena that may exacerbate pollution in such online environments (e.g., facilitate the spread of misinformation). Therefore, to achieve a human-centered AI that positively impacts society, it is necessary to understand how AI - in its different forms - can facilitate and influence emerging social behaviors. By enhancing our comprehension of how AI interacts with social phenomena, we can use it to limit negative consequences and promote favorable results supporting social well-being.

**Sciences of Social Phenomena.** How can science understand this? Social systems - and even more so socio-technical systems - are complex and need a complexity approach. The study of social phenomena has a long history, dating back to the ideas of 18th-century philosophers such as Comte[47] and Hobbes[107], who proposed an approach akin to that of the natural sciences, with a final aim of finding the "equations" governing human behavior. The development of computational power and the availability of massive online data have given rise to real fields of research such as *sociophysics*[198] and *computational social science*[49], that apply tools from mathematics, physics, and computer science to the study of human behavior. With the rise of online social media, an increasing number of human interactions have left a massive digital footprint that can be exploited to study, among other things, opinion formation and diffusion dynamics. Nevertheless, in the past, social scientists primarily sought to comprehend the behavior of groups of individuals by examining the basic characteristics of individuals. However, when the objective is comprehending emergent phenomena in society as a whole, such as forming a consensus on a particular subject or the dominance of a language, it is insufficient to focus solely on individuals to grasp them. The key lies in analyzing the *interactions* between "units", e.g., people. The field of social network analysis has significantly advanced the understanding of these dynamics by employing tools from graph theory, network theory, and complex systems theory. This interdisciplinary perspective is essential for comprehending soci(o-technic)al systems and developing effective strategies to address challenges related to online social networks and human behavior in the digital age.

[2]https://bit.ly/2XyVGzE

**Opinion Dynamics modeling.** As we stated at the beginning of this introduction, understanding opinion formation and evolution is one of these challenges. We need to interweave knowledge on different levels to understand this phenomenon truly. It is not enough to understand how people – or users in the case of online environments – create lasting relationships with each other (forming social networks). We must also understand how they interact, creating complex networks of interactions that evolve over time and can incorporate interactions that go beyond the concept of a couple. Finally, it is essential to understand what individual factors play a crucial role in the process and how these interact within the network. Our unit of analysis, in this case, needs attributes to enrich its meaning, which may change over time due to relationships and interactions. The quest to grasp the fundamental mechanisms of such a complex phenomenon has led to a significant body of literature on opinion dynamics models, which serve as a primary means for understanding the emergence of various phenomena at the opinion and public level, such as consensus formation or polarization. Such models generally consider a population of individuals and numerically simulate the interactions between them, or whenever possible, they compute the final state analytically. Such processes are normally governed by rules - often even very simple equations - developed according to empirically observed sociological behaviors, chosen by the scientists (to try) to reproduce patterns observed in the real world and provide a causal explanation of them.


FIGURE 1.2: Positive feedback loop to understanding online opinion evolution

**The goal.** Within this very open area of research, this thesis aims to understand how the interplay of biases and network effects drives the process of opinion formation and diffusion in and through online social networks, building on opinion dynamics literature with a social network analysis lens.

To tackle this goal, our approach was two-fold:

A we developed models of opinion dynamics that assume specific characteristics of the process of opinion formation and simulate the consequences in the long term;

B we used a hybrid approach assuming an existing opinion dynamics model and using it to estimate the level of "open-mindedness" within real online discussions;

We finally closed the feedback loop by using the estimated values from real data to simulate a model to check which conditions had made the actual result more likely and "validate" the theoretical conclusions previously developed.

**Thesis Structure.** To coherently organize this work with respect to such points, we divide the thesis into three macro parts as in the following:

PART I: State of the art.

This part introduces the main opinion dynamics models employed as baselines in the present thesis. Specifically, in Chapter 2, we introduce the study of societies through quantitative tools – which is now object of the field of sociophysics and computational social sciences – focusing in particular on the structures used to model societies (in Section 2.1), e.g. (dynamic) networks and higher order structures and we briefly overview the main approaches to model processes on networks (in Section 2.2); in Chapter 3, we delve into the milestones of opinion dynamics, giving in particular more in-depth details on the Deffuant-Weisbuch model [54] and its major extensions, which form a starting point for several works in this thesis (see Section 3.2.4); finally Sections 3.3 and 3.4 and Chapter 4 give a – non exhaustive – overview of the most recent advances in the field of opinion dynamics focusing respectively on analyzing the impact of different underlying structures and characteristic elements of the digital space, e.g. algorithmic bias. The chapter ends with a small survey of data sources and approaches to bridge the gap between models and data in approaching such biased realities (see Section 4.2)).

PART II: Models for Biased Digital Environments.

Opinion dynamics modeling requires a deeper study of models that better reflect reality. In particular, this thesis focuses on modeling online discussions – with their specific characteristics detailed already in Chapter 1. In Chapter 5, we studied the impact of external information sources in a biased environment. In Chapters 6 and 7, on the other hand, we explored the impact that different underlying structures can have, respectively, by studying different network topologies, adaptive systems, and higher-order structures. All these models align with this thesis's scope: improve the state-of-the-art of opinion dynamics, with a focus on digital realms and their characteristic pollution elements.

PART III: Applying models to data: hybrid approaches to analyze Polluted Information Environments.

In this part, we developed methodologies to exploit opinion dynamics models to extract knowledge from real data. In particular, in Chapter 8 we outline the methods for estimating open-mindedness on both networks (Section 8.1.1) and hypergraphs (Section 8.2.1) and we present applications of these methodologies to political discussions on Reddit during Trump's presidency (Section 8.1), polarized debates on Reddit (Section 8.2), and controversial discussions on Twitter (Section 8.3).

Some of the chapters introduced in this thesis have already been presented at conferences and/or published in journals:

- Chapter 5 and Section 8.3 that introduce the extension of [204] to account for mass media influence and test the extended model against a real discussion on Twitter are based on the following published and submitted work:
  - Pansanella, V., Sirbu, A., Kertész, J., Rossetti, G. Scientific Reports (2023). Mass Media Impact on Opinion Evolution in Biased Digital Environments: a Bounded Confidence Model.

- Chapter 6 that studies the opinion dynamics model developed in [204] on complex network structures is based on the following published work:
  - Pansanella, V., Rossetti, G., Milli, L. (2022). From mean-field to complex topologies: network effects on the algorithmic bias model. In Complex Networks And Their Applications X: Volume 2, Proceedings of the Tenth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2021 10 (pp. 329-340). Springer International Publishing.
- Chapter 7 that extends the model in [204] to account for adaptive networks topologies and higher-order interactions is based on the following published work:
  - Pansanella, V., Rossetti, G., Milli, L. (2022). Modeling algorithmic bias: simplicial complexes and evolving network topologies. Applied Network Science, 7(1), 57.
- Chapter 8 that describe methodologies to estimate the user-level open-mindedness from a real online discussion and three case studies on Reddit and Twitter are based on the following published article and student thesis:
  - Pansanella, V., Morini, V., Squartini, T., Rossetti, G. (2022, November). Change my mind: Data-driven estimate of open-mindedness from political discussions. In International Conference on Complex Networks and Their Applications (pp. 86-97). Cham: Springer International Publishing.
  - Ferro, G., Pansanella, V., Rossetti, G. (2023, April). Modeling peer pressure: a data-driven estimate of open-mindedness from high-order online political discussions. University of Pisa, Italy. MSc Data Science and Business Informatics.

Some works include opinion dynamics models that have been implemented for research dissemination and open-source research. An online implementation is available in the NDlib Python library for all these models: https://github.com/GiulioRossetti/ndlib.git.

# Part I

# The Physics of Societies: Studying Human Behavior with Quantitative Tools

# Chapter 2

# The Structures of Societies

The field of **Computational Social Science** [144] is a rapidly evolving domain characterized by the application of quantitative methodologies to **study the complexities of social systems**.

This Chapter and the following two aim to provide a brief overview of the state of the art in this interdisciplinary field, focusing specifically on the tools and techniques that underpin this thesis work.

The complexity of human behavior – and the multitude of factors influencing it – make it a challenging subject to study. For instance, a person's political orientation is shaped by a myriad of elements, including their upbringing, socio-economic context, personal bias, and the information they consume. Predicting an individual's future opinion on a single issue may be impossible and unnecessary.

However, when the goal shifts to predicting the result of an election, the law of large numbers averages over individual fluctuations, and general trends emerge. Societies are – in fact – more than the simple sum of their parts, constituting a proper *complex system* and requiring a complex system approach to understand its regularities and emergent behavior, as it is already done in the fields of ecology, neurology and economics.

Despite their differences, all complex systems exhibit similar emergent behaviors at the aggregate level, akin to those observed in statistical physics with gases. Even if humans are more complicated than gas molecules, examples of emerging regularities can be seen in time cycles in transport [16], order/disorder phases for segregation [197], culture [9], language [174], and city structure and size [19]. Such regularities "emerge" from the interaction of a large number of components of such systems (and are often referred to as emerging behavior).

The quantitative approach to studying human and societal behaviors has a rich history, tracing back to the 17-18th century. Hobbes, in 1693, was struck by the work of Galileo on motion and elaborated on the idea of representing society in terms of the laws of motion [107]. In *A Treatise of Human Nature* [118], Hume proposed a new science of man in the spirit of mathematics and physics. Comte then coined the term **Social Physics** [47] as the science that studies the social phenomena as subject to natural and invariable laws (then varied it to **Sociology** when he found out that Belgian statistiscian Adolphe Quetelet too wrote an *Essay on Social Physics*). In light of his positivist philosophy, he emphasized the importance of empirical observation and measurement for understanding social phenomena, a task that was arduous during the 18th century. Despite the critics, this idea made it to the digital age [182], where our everyday actions leave a massive amount of digital traces that can be studied and where advanced computational techniques to analyze them are available. Using a quantitative approach, hidden patterns can be uncovered, emergent phenomena identified, and predictions about human behavior at both the individual and collective levels can be made. Following the lead of physics, mathematics,

and computer science, scholars have finally been able to use quantitative methodologies to study the complexities of social systems. The field of **computational social science** has emerged in recent years as an interdisciplinary domain that combines insights from sociology, psychology, physics, and computer science to uncover patterns, dynamics, and underlying mechanisms that govern human behavior [144, 49].

To better understand complex systems, it is important to recognize that they are underpinned by networks that delineate the interactions between their individual components. This is particularly relevant when studying societies as complex systems, as understanding the underlying networks is crucial to comprehending societal behavior. Therefore, in order to effectively apply quantitative tools to the study of societies, it is necessary to first gain a solid understanding of the field of network science and how it can be used to map and analyze the networks that underlie complex systems. The field of network science is used today to understand diverse phenomena, and network scientists have been able to uncover universal properties of complex social networks.

Before digging into the details of network theory, let us outline some basic **definitions** common to every framework introduced in the present work:

- **Unit, element, node, vertex**: an individual object, agent, part of the system. In the present work, we normally denote a set of $N$ nodes as $V = \{v_1, v_2, ..., v_N\}$.
- **Interaction, relation**: a set $\mathcal{I} = [v_0, v_1, ..., v_{k-1}] \in V$. We denote the set of interactions as $R$
- **Property, attribute**: information attached to a node or relation. We call the set of properties $A$ and let $a$ be the assignment map sending $V \times R \to A$
- **System**: a collection of units $V$, relations $R$ and attributes $A$, such that the collection needs no other pieces in order to function completely or to interact autonomously with its environment.

In the forthcoming discussion, we will outline three different frameworks to represent complex systems: graphs, temporal graphs and higher-order structures. The graph framework (cf. Section 2.1.1) is the predominant tool in complex network analytics. We will briefly review its primary topological characteristics and the contributions it has made to the field of network science. Conversely, the temporal (or dynamic) framework (cf. Section 2.1.3) and the higher-order framework (cf. Section 2.1.2) – both increasingly present in the study of complex phenomena – will be overviewed in their most basic characteristics necessary to understand the current thesis works. We conclude the present chapter with a discussion on processes unfolding upon an underlying structure (cf. Section 2.2) – that can be modelled chosing one of these frameworks. This discussion is concluded with a mention to phenomena that emerge from the interplay of structural and behavioral dynamics, particularly important for the present work.

The understanding of the tools of network science will be later necessary to understand the next three Chapters, in which we will mainly discuss models of human behavior. Specifically, in Chapter 3 we will delve in mathematical models of opinion formation covering the milestones and moving to more complex models in Sections 3.3 and 3.4 and finally – in Chapter 4 – discussing the main approaches to the study of opinions in the digital era from a computational social science approach.

## 2.1 From Graphs to Higher Order Systems: Modeling Frameworks for Societies' Structures

### 2.1.1 Graphs

Complex networks provide a robust model for describing a range of phenomena, with graphs offering a powerful representation for encoding complexity in a mathematical framework.

We will analyze the essential characteristics and properties researchers require to comprehend complex systems. Specifically, we will examine the importance of graph characteristics.

**Basic definitions** A graph is a mathematical structure used to model **pairwise** relationships between objects.

**Definition 1 (Graph)** *A graph $\mathcal{G}$ is formally defined as a pair $\mathcal{G} = (V, E)$ comprising a set $V$ of vertices (also called nodes) and a set $E \subseteq V \times V$ of edges (also called links).*

An edge $e = (u, v)$ is said to join the vertices $u$ and $v$ and to be incident on $u$ and $v$. A vertex may exist in isolation, i.e., it may not be joined to any other vertex. Graphs can be either directed or undirected. In an undirected graph, the edges are unordered pairs of vertices (i.e., $(u, v) \in E$ iff $(v, u) \in E$), whereas in a directed graph, the edges are ordered pairs of vertices.

**Degree, degree distribution, scale-freeness** One key property of a node is the number of other nodes it is connected to. Therefore, we define the degree of a node as:

**Definition 2 (Degree)** *The degree of a node $v$, $k_v$, is the number of edges incident on node $v$.*

In the case of directed graphs, it is necessary to differentiate between the in-degree (in-degree) $k_v^{in}$ and out-degree (out-degree), $k_v^{out}$. The total degree of the node is obtained by adding these two, $k_v = k_v^{in} + k_v^{out}$. The total number of edges in the graph can be determined from the degree of each node. Specifically, Equation 2.1 gives us the total number of edges as half the sum of node degrees, where $k_v$ is the degree of node $v$. Equations 2.2 indicate that the total number of edges equals the sum of in-degree and out-degree for each node, where $k_v^{in}$ and $k_v^{out}$ are the in-degree and out-degree of node $v$, respectively.

$$L = \frac{1}{2} \sum_{i=0}^{N} k_{v_i} \tag{2.1}$$

$$L = \sum_{i=0}^{N} k_{v_i}^{in} = \sum_{i=0}^{N} k_v^{out} \tag{2.2}$$

The connectivity of the network is indicated by the average degree, which can also be expressed in terms of the total number of nodes $N$ and the total number of edges $L$ as:

$$\langle k \rangle = \frac{1}{N} \sum_{i=0}^{N} k_{v_i} = \frac{2L}{N} \tag{2.3}$$

for undirected graphs and

$$\left\langle k_i^{in} \right\rangle = \frac{1}{N}\sum_{i=0}^{N} k_i^{in} = \langle k_i^{out} \rangle = \frac{1}{N}\sum_{i=0}^{N} k_i^{out} = \frac{L}{N} \tag{2.4}$$

$$p_k = \frac{N_k}{N} \tag{2.5}$$

for directed ones.

It is also possible to define the degree of a node by using neighbor sets. Specifically, the neighbor set of a node $u$ is defined as $\Gamma(u) = \{v \in V | (u,v) \in Eor(v,u) \in E\}$. Using this definition, $u$'s neighbor set's cardinality, i.e., $|\Gamma(u)|$, represents $u$'s degree.

We defined $\langle k \rangle$ as the graph's average degree, useful for describing the system through a global property. However, real-world networks frequently exhibit heterogeneous behaviors that averaged measures like $\langle k \rangle$ cannot capture.

In networks where the degree distribution follows a power law, indicating that there are numerous nodes with low degrees and a few nodes with very high degrees, the average degree may not accurately represent the network's characteristics because it fails to consider the presence of these highly connected nodes, also known as "hubs". These networks are referred to as "scale-free" (first defined in [15], i.e., the network does not have a characteristic scale of "size".

On the other hand, analyzing the degree distribution can provide valuable insights into the structure and behavior of the network. The presence of hubs can significantly impact the network's resilience to attacks or failures. Removing a hub can have a much more substantial effect on the network's connectivity than removing a node with a low degree. Inevitably, hubs can also lead to imbalances in resource allocation within a system. This is evident in various contexts, such as the uneven popularity distribution on social media, disparities in market success, or even in protein-to-protein interaction networks. In socio-economic contexts, these imbalances can result in extreme inequalities that are unsustainable for the societal goals we aspire to achieve. Therefore, when studying networks with scale-free properties, it is crucial to consider the degree distribution to understand their characteristics comprehensively.

More details on degree distributions will be discussed when introducing different network models in Section 2.1.1.

**Paths, clustering coefficient, small-worldness** One of the main roles networks play is connecting the local and the global, explaining how simple processes at the level of individual nodes or links can have complex effects that spread across the population. Beyond global network connectivity, of which degree is a key indicator, other properties are necessary to understand network dynamics.

Given two nodes $u, v \in V$, a **walk** between them is defined as the sequence of edges crossed during a visit starting at $u$ and ending at $v$. In complex network tasks, paths are important to quantify coupling nodes' distance. A walk commonly has a length, defined as the number of edges between the starting and the ending point. A walk with the minimum possible length is called a shortest path, and it identifies a path with the minimum number of edges between two coupling nodes. The length of a shortest path from $u$ to $v$ represents the distance $d(u,v)$ between these vertices. The diameter $diam(\mathcal{G})$ of a graph $\mathcal{G}$ is the maximum of the distances between nodes in $\mathcal{G}$, i.e., the length of the longest shortest path between any two vertices. Similarly

to the previous overview about the degree, real-world systems' behavior of shortest paths and diameters unveils not trivial patterns. The popular experiment of Milgram about the *six degrees of separation* [159] demonstrated that social networks are *small world*, namely, they tend to reduce distances between nodes. A useful quantity to describe this property is the average path length $\langle d \rangle$ of a graph, namely the average of the shortest paths for each pair of nodes. In real-world networks – particularly, in social networks – , we call *small world* the property that $\langle d \rangle$ depends logarithmically on the system size [14]: $\langle d \rangle = \frac{logV}{log\langle k \rangle}$, that is to say that the denser is the network, the smaller will be the distance between the nodes.

To better explain the small-world effect in complex systems, we also need to introduce the notion of the clustering coefficient. The clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together by forming triangles. In other words, it answers the following question: "What fraction of my neighbors are connected?" The clustering coefficient of a node $u$ computes the ratio of closed triangles over all the triplets of $u$'s neighbors: $C(u) = \frac{2e_u}{k(k-1)}, \in [0, 1]$, where $e_u$ indicates the number of triangles formed by $u$'s neighbors, and $k$ is $u$'s degree.

The clustering coefficient is particularly useful to capture the triadic closure effect: if a connection between A and B and between B and C exist, there is a strong tendency to form a connection between A and C. The cohesive power of such connections, also referred to as *weak ties* [97], are shown to be important in social networks for many reasons, from the diffusion of influence and information to community organization. With a lower value of the average shortest paths of a graph, high clustering coefficient values tend to be observed in social networks and many other complex systems, such as power grids and brain networks [226].

#### Models of complex networks

The previously introduced metrics are not just useful to analyze real networks, but also to create "random" networks incorporating some key properties observed in the real world. Such models are commonly exploited for different tasks. Firstly, to better understand a common property that emerges in real networks by generating graphs with simple rules. Secondly, to test algorithms or simulate dynamical processes over random networks while maintaining certain properties of real networks. Statistical methods are used to match key network metrics between synthetic and real networks to ensure accuracy. Finally, to determine the statistical significance of observations, multiple random networks with the same characteristics as the original network are generated and compared. Now let's examine some of the main models, specifically focusing on the ones employed in Part II.

**Complete network.** A complete graph is a simple graph in which an arc connects each pair of vertices. If the network has $N$ nodes, the total number of links present is equal to the maximum number of possible links $L = L_{max}$ Obviously, in a complete graph, the degree of each node is equal to the average degree, which is equal to $N - 1$. A complete graph is often called a clique. It is clear that, in reality, there are few complete networks. Nevertheless, this approximation is often used in diffusion models and other models of social behaviors (see homogeneous mixing hypothesis in Section 2.2). In reality, the number of nodes $N$ and links $L$ can vary enormously and in most cases $L \ll L_{max}$, reflecting the fact that most real networks are "sparse" that is, the number of links present is only a small fraction of the number expected for a complete network with the same number of nodes.

**Random networks.** Real-world networks may initially appear to be randomly connected. The Erdős-Rényi random network model, introduced by mathematicians Paul Erdős and Alfred Rényi in 1959 [64], generates random graphs with distinct features. While in social networks new connections depend on existing ones, in random graphs connections are formed without any prior knowledge of those already existing. In this context, therefore, the term "random" is used to indicate "statistical independence". There are two variants of the Erdős-Rényi model: the $G(N, L)$ model, which generates a random network with $N$ nodes and $L$ random links, and the $G(N, p)$ model, which generates a network where each pair of nodes is connected with probability $p$. The $G(N, p)$ model is more commonly used, as it allows for the computation of many important network properties. The expected number of links in a random network is $\langle L \rangle = p\frac{N(N-1)}{2}$, with an average degree of $\langle k \rangle = p(N-1)$. When the network is sparse, the degree distribution is approximated by a Poisson distribution. Random networks exhibit the "small-world" property, and the average degree $\langle k \rangle$ grows logarithmically with $N$. However, the average clustering coefficient in real-world networks is higher than predicted by random models.

**Scale-free networks.** A scale-free network is a network whose degree distribution follows a power law:

$$p_k \sim k^{-\gamma} \tag{2.6}$$

The **Barabasi-Albert model** provides a mathematical framework to understand and generate scale-free networks, shedding light on the underlying mechanisms that drive their formation. It was introduced by Albert-László Barabási and Réka Albert in 1999 [15], and it generates a scale-free network by following a *preferential attachment* mechanism. This mechanism is also known as "rich gets richer" or "the Matthew effect". The basic idea is that new nodes entering the network are more likely to connect to nodes that already have a high number of connections. This preferential attachment process leads to the formation of hubs, which are highly connected nodes. The Barabasi-Albert model works like this: starting with $m_0$ linked nodes, each time step a node $u$ is added with $m < m_0$ links, following:

$$P_{uv} = \frac{k_v}{\sum_j k_j} \tag{2.7}$$

After $t$ steps, the BA model produces a network with $N = t + m_0$ nodes and $m_0 + m_t$ edges, characterized by a power-law degree distribution with $\gamma = 3$.

For a comprehensive study on graph properties and network models, refer to [14, 50].

**Networks with fixed degree distributions.** To achieve a broad degree distribution – instead of employing a generative process – , the simplest approach is to enforce it directly. This involves establishing a degree sequence – using it as an input into the network generation process – and assigning each node a degree value from the sequence. This straightforward method forms the foundation of the **configuration model** [169], requiring no complex mechanics. The model generates a network where each node has a predetermined degree $k_i$, while the connections between nodes are formed randomly. By applying this procedure repeatedly to the same degree sequence, we can create multiple networks with the same degree distribution $p_k$. By knowing the number of nodes with a given number of edges, we

can disregard the actual connections and instead focus on connecting the "stubs" (half-links) of each node. This approach allows us to create a network matching the desired degree distribution while retaining flexibility in forming node connections. The probability of having a link between nodes of degree $k_i$ and $k_j$ is:

$$p_{ij} = \frac{k_i k_j}{2L - 1} \tag{2.8}$$

Indeed, a stub starting from node $i$ can connect to $2L - 1$ other stubs. Of these, $k_j$ are attached to node $j$. So, the probability that a particular stub is connected to a stub of node $j$ is $k_j/(2L - 1)$. As node $i$ has $k_i$ stubs, it has $k_j$ attempts to link to $j$, resulting in eq. 2.8.

**Networks with mesoscale structure** The configuration model can incorporate high clustering while still randomly closing triangles, resulting in a random distribution of triangles within the network. However, real-world networks often exhibit higher triangle correlation and form modules and assortative communities, presenting a meso-scale topology hidden within the complex network structure.

In general, we define a **community** as a subset of nodes that are more interconnected with each other than with the rest of the network. The primary objective of community detection algorithms is to identify such meso-scale topologies. However, different algorithms have been developed to generate networks with a given meso-scale structure, which can be used to test the communities found in real-world networks against them. These benchmarks can also be used as generative network models to create networks with a given community structure, which can serve as a realistic proxy for a real-world network to test algorithms and simulate processes on.

Among the most famous generators used for classic community discovery, we find the Girvan-Newman (GN) [93] and the Lancichinetti-Fortunato-Radicchi (LFR) [140] benchmarks, as well as the family of stochastic blockmodels (SBMs) [109, 131]. The GN benchmark [93] generates networks with a given community structure by starting from equally sized communities with a given degree distribution and with a mixing parameter $\mu$ indicates the share of edges that a node has to nodes that are not part of its community. This benchmark has many limitations since it has a fixed number of nodes, its degree distribution is binomial, and it generates equally sized communities, all characteristics rarely encountered in real networks. The LFR benchmark [140] fixes all the previously stated limitations, allowing for a user-defined number of nodes, and distributes both node degrees and community sizes according to a power law. The LFR benchmark has many parameters to generate realistic networks: (a) $\alpha$, the exponent of the power-law degree distribution of the graph, (b) $\beta$, the exponent of the power-law size distribution of the communities in the graph, (c) $|V|$ number of nodes in the graph, (d) $\langle k \rangle$ average degree of the nodes, (e) $s_min$ and $s_max$ as the minimum and maximum community sizes and (f) finally $\mu$ the mixing parameter regulating the fraction of edges going outside their planted communities. In Chapter 6 we employed this benchmark to generate synthetic networks with a given community structure and studied the unfolding of an opinion dynamics model on these, under different initial conditions. The different underlying structure impacted the dynamics as we will see in Chapter 6. In the SBM [109], nodes are assigned to one of $k$ user-defined communities; then, the links are placed

independently between nodes with probabilities that are a function of the community membership of the nodes. The classical SBM fails to reproduce the degree distribution of real networks (power-law/scale-free): a degree-corrected SBM allows the identification of heterogeneous node degrees [131].

In the present section we just scratched the surface of graph theory, focusing on the notions necessary to understand the subsequent chapters. Despite its proven usefulness, however, such a framework poses some limitations that we will address in the next section, going beyond the notion of pairwise interactions.

### 2.1.2 Higher-order structures

Like all models, graphs are simply representations we use to understand real-world phenomena. However, there are instances where graphs may not be expressive enough to describe specific patterns, limiting their usefulness. The limitation of networks is that they only capture pairwise interactions, whereas many systems exhibit group interactions. In social systems [190], ecology [145], and biology [58], among other examples, connections and relationships often occur between groups of nodes rather than pairs.

Initial research on networks has examined collective interactions, but often interpreting them through the lens of pairwise networks. This approach assumes that the influence among elements in a complex system can be broken down into individual, two-way connections.

Recent successful research in network science, such as complex contagion or majority-rule opinion dynamics models, accounts for multiple simultaneous interactions. However, it is worth exploring mathematical models that can explicitly and naturally describe group interactions, generalizing pairwise links to arbitrary node sets. Simplicial complexes and hypergraphs are the ideal options to offer such descriptions.

**Interactions** We must first define an interaction before digging into the mathematical structures used to represent higher-order interactions.

**Definition 3 (Interaction)** *An interaction is a set $\mathcal{I} = [v_0, v_1, ..., v_{k-1}]$ containing an arbitrary number of k basic elements of the system under study (i.e., nodes or vertices).*

To resume the examples from the previous paragraph, an interaction $\mathcal{I}$ can describe the coauthors of a scientific paper, a set of users in a thread of comments in an online discussion, or a group of people visiting the same location at a given time. The *order* of an interaction involving $k$ nodes is $k-1$. So a $0-$order interaction is a node, a $1-$order interaction is an edge (as defined in 1), and a $2-$order interaction involves three nodes. A higher-order interaction is an interaction with $k \geq 2$, and a *Higher-Order system* displays higher-order interactions (conversely, a low-order system is a system where only nodes and edges are present).

#### Simplicial Complexes

Simplicial Complexes (SC) are mathematical structures developed in algebraic topology that capture higher-order interactions between constituents of complex systems. SC have become increasingly applicable to real data due to a growing computational toolset, becoming increasingly popular for the representation of social systems [187], for example, in the context of diffusion analysis, e.g., studying social contagion with

simplicial complexes [119], in time-varying settings as well [40]. They have also been used to study the network structure of scientific revolutions [129] and the evolution of higher-order linguistic networks in scientific texts [41]. This work has exploited SC in Chapter 7 to model group dynamics and peer pressure influencing opinion change.

Just like graphs are collections of edges, Simplicial Complexes are collections of simplices $\sigma$.

**Definition 4 (Simplicial Complex)** *A Simplicial Complex $\mathcal{C} = (V, S)$ is defined as an ordered pair of sets, where V are the vertices and S are the simplices, each of which is a finite subset of V, subject to the requirement that if $\sigma \in S$, then every subset $\tau$ of $\sigma$ is also in S. Subsets $\tau \in \sigma$ are faces of $\sigma$.*

From the definition, it follows that a collection of $n$ simplices $K = \{\sigma_1, \sigma_2, ..., \sigma_n\}$ is a valid simplicial complex if for every k-simplex $\sigma = [v_0, v1, ..., v_{k-1}] \in K$ all its subfaces of any dimensions belong to $K$, too.

For example, if the triangle [a, b, c] $\in K$, then we also require [a],[b],[c],[a, b],[a, c],[b, c] to belong to $K$.

As defined, a simplicion is a clique that guarantees all induced sub-cliques exist (while in a graph/clique complex, cliques make it impossible to distinguish the two cases in which a sub-face is present or not).

Because any given simplex must "contain all of its faces", it suffices to specify only the maximal simplices, those which do not appear as faces of another simplex. This dramatically reduces the amount of data necessary to specify a simplicial complex, which helps make both conceptual work and computations feasible.

Although simplicial complexes overcome some of the problems encountered by other lower dimensional representations, they are still quite limited by the requirement on the existence of all subfaces.

This requirement, i.e., the existence of all subfaces, is stronger with respect to what is required by the hypergraph representation (see sec. 2.1.2): subfaces information might often be unavailable from real-world networks and applied tasks.

### Hypergraphs

Hypergraphs offer the broadest and most unrestricted representation of higher-order interactions. In technical terms,

**Definition 5 (Hypergraph)** *A hypergraph $\mathcal{H} = (V, H)$ is a mathematical structure defined by a pair $(V, H)$, where V is a non-empty set of nodes and H is a set of non-empty subsets of V, called hyperedges.*

Each hyperedge in $H$ represents a higher-order interaction among the nodes in $V$. Differently from simplicial complexes, hypergraphs can include a 2-order interaction [a, b, c] without any requirement on the existence of the 1-order interactions [a,b], [b,c], and [c,d]. We exploited this higher versatility to model online discussions in Chapter 8 and Section 8.2.

**Incidence matrix** An incidence matrix is a mathematical object that describes the relationship between two classes of objects.

**Definition 6 (Incidence Matrix)** *An incidence matrix $I = \{I_{i\alpha}\}$ is an $n \times m$ matrix where n is the number of nodes and m is the number of relationships. The entry $I_{i\alpha}$ in row i and column $\alpha$ is $w > 0$ if node i and relationship $\alpha$ are incident, and zero otherwise.*

In the case of a graph, $m$ will be the number of edges, and in the case of a hypergraph, $m$ will be the number of hyperedges. If each node can be present in each hyperedge only once, the matrix entries can be either 1 or 0, otherwise, any positive value represents the times the node appears in the hyperedge. Notice that the incidence matrix can also be seen as the adjacency matrix of a bipartite graph with two node sets, one of size $n$ and one of size $m$.

In the case of simplicial complexes, the incidence matrix between nodes and simplices can be defined in the same way.

From the incidence matrix, we can compute the node degree.

**Definition 7 (Node Degree)** *In a hypergraph, the degree of a node v is defined as the number of hyper-edges that contain v. The degree of node i is the sum of the elements of the ith-row of the incidence matrix, i.e., $k_i = \sum_h I(i,h)$, where h ranges over all hyperedges in the hypergraph.*

In a graph, the column of an incidence matrix always sums to 2 as the relationships described are always between two nodes of the graph. In a hypergraph (or simplicial complex), the rows of the matrix can have more than two non-zero elements as each hyperedge (simplex) can describe interactions among more than two vertices. The degree of a node in a hypergraph can be seen as a measure of its involvement in higher-order interactions.

In a hypergraph, the size of a hyperedge $h$ is defined as the number of nodes that participate in the interaction represented by $h$. Formally

**Definition 8 (Size)** *The size of a hyper-edge h is given by the sum of the entries in the column of the incidence matrix I that corresponds to h: $|h| = \sum_v I(v,h)$ where v ranges over all nodes in the hypergraph.*

The size of a hyper-edge is a measure of the number of nodes that are involved in a higher-order interaction.

**Adjacency Matrix** From the incidence matrix of a graph, we can also construct another matrix that fully encodes the connectivity of the graph, the adjacency matrix $A$.

**Definition 9 (Adjacency Matrix)** *The Adjacency Matrix $A = IWI^T + D$ is a $n \times n$ matrix where for $i = j$, $a_{ii} = k_i$ (the number of hyperedges node i belongs to), while for $i \neq j$, $a_{ij} = w$ iff nodes i and j are adjacent, that is, they appear together in one, or more, hyperedges. Here w is the number of hyperedges containing i and j.*

### 2.1.3 Dynamic networks

While enriching the representation of complex systems with the concept of polyadic interactions, higher-order structures – as well as the low-order counterparts – miss a fundamental element to understanding real systems: **time**.

Most real-world networks are dynamic in nature, with nodes and edges appearing and disappearing over time, and the nature of relationships changing as well [112].

For the sake of completeness, we will surface an introduction to temporal (or dynamic) networks in the following section. This framework has been only marginally employed in this thesis work: in Chapter 7, the concept of adaptive topology is present, which can be better understood in the light of certain concepts pertaining to

temporal networks; furthermore, in Chapter 8, the analyzed online discussions are modeled as dynamic networks.

Examples of such networks include online social networks, face-to-face contact networks, infrastructure networks, biological networks, brain/neural networks, and ecological networks. To analyze these systems effectively, adopting a time-aware network representation is crucial.

While analyzing such systems, the benefit of a time-aware network representation is incalculable.

To move to a temporal representation, we must first differentiate between the types of tie evolution we are dealing with. On the one hand, we can have stable connections that involve long-term or short-term *relations*, such as friendships, work colleagues, family members, project collaborators, or paper co-authors. On the other hand, we can have unstable connections that involve node *interactions* that can be instantaneous (e.g., an email or a text message) or have a certain duration (e.g., a phone call or a face-to-face contact) [112].

The different time scales of the topology evolution call for different representations. However, both interactions and relations can be represented through graph series.

**Definition 10 (Graph Series)** *A graph series* $DN = [G_1, G_2, ..., G_k]$ *is a sequence of k graphs* $G_i = (V_i, E_i)$

Graph series can be obtained as a series of *snapshots*, each one of them corresponding either to the state of the network at a time $t$ (relation network); otherwise, if we are dealing with interaction networks, each individual snapshot is constructed by aggregating the interactions that occurred within a certain time frame or time window. It is worth noting that, in the aggregation process, time is divided into windows, and for each window, all node pairs with at least one connection occurring within that segment are included in the resulting snapshot graph. Consequently, the order of the connections within the time window, but the snapshot representation is considered stable. However, the selection of an appropriate time window is not a trivial task. The choice of a wider time window may seem to guarantee greater stability, but recent literature has shown that this assumption is not always accurate and has quantified the potential discrepancies that may result from relying exclusively on wider windows for stability. In the present work, discussions on Reddit and Twitter will be modeled as graph series, as we will see in Sections 7.3, 8.1 and 8.3.

In this section, we briefly introduced dynamic topologies. However, there is another type of dynamics that occurs on networks, such as diffusion phenomena (epidemics, information, innovations, opinions, etc.). In the next section, we will give an overview of the processes that unfold on networks before moving on to a detailed review of the literature on opinion dynamics, which is the core of this thesis.

## 2.2 Social Dynamics: Processes Unfolding on (Network) Structures

Collective phenomena emerge from the interactions of (individuals as) elementary units in complex (social) systems.

In Section 2.1, we underlined how nontrivial topological structures emerge from the self-organization of human agents in social (network) structures.

Nevertheless, such structures also serve as substrates of (social) dynamics, and unsurprisingly, topological characteristics play a pivotal role in the unfolding of the

dynamics. Even if the evolution of the process occurs independently of the structure, i.e., we consider a static structure, the two levels are interrelated.

### 2.2.1 Modeling choices for underlying structures

Understanding social dynamics requires a deep comprehension of population contact patterns, which influence the ability of system units to change states. Modeling these patterns is key to understanding how phenomena spread and can be controlled. Traditional models often use the homogeneous mixing hypothesis, assuming an equal probability of contact between any two individuals. This approach, while analytically convenient, lacks realism in social contexts. More sophisticated models incorporate social structure, classifying individuals based on demographic information. However, these models still lack a network structure. Contact network models use network structures to represent social interactions, providing potential propagation paths and influencing dynamics with their topological properties. Multi-scale models capture phenomena occurring on various spatial and temporal scales, such as disease spread, where homogeneous mixing can be assumed at lower scales. Agent-based models simulate individual agent actions and interactions in detail, assessing their system-wide effects and addressing the social system's emergence from the micro to the macro level. Accounting for a more complex underlying structure inevitably influences spreading speed and patterns, which depend on initial infection seeds, surrounding topology, node homophily degree, and other topological characteristics like scale-freeness or small-worldness.

### 2.2.2 Different dynamics require different modeling choices

From the diffusion of viruses to the emergence of cognition, from the formation of friend groups to financial crises and power outages, emergent phenomena are determined by processes that operate on an underlying complex structure. Besides tending to reduce the variability of the initial state of the system, these processes present their own peculiarities and are often studied by different lines of research.

**Disease Spreading Models and Their Influence on Social Dynamics.** The most studied examples of processes on network structures are models of disease spread in populations. In these models, nodes (representing individuals) can have different statuses, such as susceptible, infected, or immune. Mathematical models in epidemiology, such as the Susceptible-Infectious-Recovered (SIR) and Susceptible-Infectious-Susceptible (SIS) models, simplify the progression of infectious diseases. These models use differential equations with parameters representing various factors, including the number of susceptible, infectious, and recovered individuals, birth rate, natural death rate, recovery rate, and force of infection. The force of infection, which is the rate at which susceptible individuals become infected, depends on the number of infectious individuals and represents the transmission of infection. This leads to a nonlinear transmission of infection, generating rich dynamical behaviors. These models have been studied mainly with a homogeneously mixed population or accounting for some social structure. However, in many situations, each person's network of contacts is significantly smaller than the total population, meaning that interactions aren't entirely random. Network-based models address this by assigning each individual a specific set of constant contacts. Studies on networked populations have highlighted differences with standard homogeneous-mixing disease models.

While the aforementioned models have been widely used to study the spread of diseases, there's a variety of other domains where they have been successfully applied. Indeed, another long tradition of modelers that have been using similar frameworks to characterize the spreading of social phenomena, such as the diffusion of rumors and fads or the adoption of novelties and technological innovations. Much early research on information diffusion, i.e., the propagation of information through a social network structure, has been based upon the analogy with the spread of diseases.

However, in all these situations, the social nature of the contacts that mediate these processes calls for ad-hoc modeling adjustments that are not present in simple disease epidemics models. Researchers have made ongoing improvements based on classical models, developing new models such as SEIR (Susceptible Exposed Infected Removed) model, S-SEIR (Single layer-SEIR), SCIR (Susceptible Contacted Infected Removed) model, irSIR (infection recovery SIR) model, FSIR (Fractional SIR) model, and ESIS (Emotional Susceptible Infected Susceptible) model.

Despite adjustments, epidemic models lack ingredients like peer pressure and social influence, which are essential in social dynamics. For example, if I am thinking about buying the new iPhone 15, I may be convinced that it is a good choice after k% of my neighbors already bought it. Threshold models better incorporate these drivers of change [96]. In these models, the role of underlying topology can have a great impact on the result of the dynamics.

The social dynamics we mentioned are relatively fast processes, which bring the population from a disordered to an ordered phase relatively fast. However, not every social dynamic is this fast, e.g., the formation of cultural groups can take months or years [9], in the same way as the evolution of a language [51]. Both of these processes clearly show different outcomes if we account for social structure [99].

In the next chapter, we will delve into a specific kind of social dynamics, i.e., opinion dynamics. Before moving further, it is already worth mentioning that in all these models, the behavior or state being transmitted is of a binary (or discrete) nature. This simplification clearly facilitates the tracking of behavioral cascades from a source (or multiple sources) to the entire system (or parts of the system). This feature serves the purpose of studying collective decision-making processes involving two options, such as voting, adoption of innovations, and binary opinion dynamics, well. However, numerous social processes are more complex and involve a continuum of options rather than just a binary set. For instance, politics is often seen as a binary choice between left and right, but in reality, it's a spectrum that includes a wide range of beliefs and ideologies, from socialism to conservatism and everything in between and requires to be modeled accordingly.

### 2.2.3 Social Dynamics on and of networks: adaptive topologies

In Section 2.1.3, we stated that structures evolve (sometimes depending on node statuses, see, for example, homophily), while, in this section, that nodes statues evolve (sometimes depending on network structure). It is clear that, in many domains, the evolution of networks and the evolution of a certain phenomenon are interdependent.

**Adaptive networks**, also known as coevolving networks, are a modeling framework where the structure of the network and the states of its nodes coevolve over time, giving rise to complex, emergent behaviors.

Despite the thematic diversification, certain dynamical phenomena repeatedly appear in adaptive networks, including the formation of complex topologies, robust dynamical self-organization, the spontaneous emergence of different classes of nodes from an initially inhomogeneous population, and complex mutual dynamics in state and topology, assessing adaptive network as a natural framework in many different applications. They are found in technical distribution networks such as power grids, the mail network, the internet, or wireless communication networks. They also appear in natural and biological distribution networks, information networks like neural or genetic networks, and social networks. In game theory, the evolution of cooperation in simple agent-based models is studied on social networks. Adaptive networks also appear in chemistry and biology, with models involving adaptive networks having a long tradition in ecological research.

A simple framework in which the dynamical interplay can be studied is offered by contact processes, which describe the transmission of some property, e.g., information, political opinion, religious belief, or epidemic infection along the network connections. For example, what if Susceptible individuals try to avoid contact with Infected ones? In [98], authors show how this simple rewiring mechanism can lead to the emergence of global structure from local rule, with the presence of isolated infected nodes and a single tightly connected cluster of susceptible.

The emergence of Echo Chambers [43] may be a clear example of the type of phenomenon that arises from the interplay between the evolution of an attribute and the evolution of a structure. Adaptive networks may play a crucial role in understanding the formation of Echo Chambers through self-organization and homophily tendencies (individuals with similar opinions or behaviors are more likely to form connections, leading to the formation of tightly-knit communities or clusters within the network) and through a feedback loop that reinforces their beliefs, strengthening their ties with like-minded individuals and sever connections with those holding opposing views, further exacerbating the Echo Chamber effect.

Adaptive networks are a promising concept for the investigation of collective phenomena in different systems. However, they also present a challenge to existing modeling approaches and analytical descriptions due to the tight coupling between local and topological degrees of freedom. We will see how adaptive networks have been used in the field of opinion dynamics in Section 3.3 and how these can be applied to model the emergence of echo chambers in Chapter 7.

# Chapter 3

# The Opinions of Societies

In Chapter 2, we introduced the idea of society as a complex system and emphasized that every complex system has an underlying network, which can be static or dynamic, dyadic or polyadic. In the last section, we briefly discussed the different phenomena that evolve on top of these structures, mentioning **opinion dynamics**, without going into particular detail. In this chapter, we will provide a general introduction to opinion dynamics models, establish common terminology, and examine the main classical models.

## 3.1 Preliminaries

Before going into the details of the different models, some basics are necessary to understand how opinions and their dynamics are represented mathematically. The elements that, at least, must be represented in such a model are opinions, interaction patterns, and time.

Opinion Dynamics models (OD models) consider a population of agents – which may be characterized with different attributes and belong to different categories with different characteristics – and a set of connections among those agents, on which influence *can* flow. Generally speaking, such a population can be modeled with a graph $\mathcal{G}$ of $N$ agents ($N$ being the size of the population), which can go from a fully connected graph to more complex structures. A connection between two or more agents can also be characterized with different attributes, e.g., defining the level of influence of that connection. In OD models, each agent $i$ holds an opinion or a set of opinions. Focusing on the single opinion case, the opinion of agent $i$ at time $t$ is a variable $x_i(t)$ which can be binary $x_i(t) \in \{0,1\}$, discrete $x_i(t) \in \{1,2,...,m\}$ or continuous (but bounded) $x_i(t) \in [0,1]$. The set of all possible opinions is called the opinion space. Let us further denote by $\mathbf{x^{(t)}} = [x_1^{(t)}, x_2^{(t)}, ..., x_N^{(t)}]$ the vector of opinions of all agents at time $t$; in some articles it is also called a profile. To define the dynamics, we need two rules: (a) a rule to choose the interacting partner(s) among the possible ones (which may incorporate parameters representing social dynamics and external factors) and (b) an update rule for the opinion of the agent(s) after an interaction (which normally incorporates a set of parameters representing psychological aspects of the opinion evolution process). Update rules can be mainly divided into two categories: (a) synchronous, where every agent updates their opinion at time $t$, given the profile of the population at time $t$, or (b) asynchronous, where agents update their opinion sequentially, through one-to-one or one-to-many interactions, and the profile of the network can be updated continuously or after a set of $K$ interactions. The dynamics end either when the population reaches an equilibrium state (convergence) or when a stopping condition is met. The equilibrium state can be

consensus, Polarization, or fragmentation. There is no agreed definition of Polarization, but normally – in OD literature – it refers to two groups of agents with opinions at opposite ends of the opinion space. However, in some works, Polarization simply indicates the existence of two groups (not necessarily distant in the opinion space).

## 3.2 Milestones

As we saw in Section 2.2, many phenomena can be modeled with the "change of state" abstraction, i.e., agents can be in a finite set of different states or – in other words – choose among a finite set of actions. It must be recognized, however, that this is not a modeling choice suitable to every context or domain; some social phenomena require a continuous approach.

In the opinion dynamics literature, when $x_i \in \{o_1, o_2, ..., o_m\}$, i.e., the agents have a finite set of possibilities, we are in the realm of discrete models. Such models can be useful to model the choice of a political candidate, a product, or answers in a survey. Most discrete models are *binary state* models, i.e. $x_i(t) \in \{0, 1\}, \{-1, 1\}$. When we only have two states, the dynamical rules of the model are incorporated in the transition rates, the probability – at time t – that an agent changes their opinion from one state to the other. In analogy with epidemic spreading (cf. Section 2.2), agents become infected with rate $F$ and recover with rate $R$. The functional form of the transition rates $F$ and $R$ determines the type of model we have as well as its stationary states of collective opinion.

In the field of physics, research has been largely focused on discrete opinion spaces due to their strong analogy with spin systems. In some cases, these spaces have been expanded to incorporate more than two spin values, resulting in a closer approximation to continuous opinion dynamics.

### 3.2.1 Voter models

In its original formulation, the *Voter model* was intended for the study of the competition of species [46]. However, it was soon employed to explain opinion formation dynamics [110]. The most general formulation in the family of *Voter models* is the Non-linear Q-voter model [35]. The basic idea is that $N$ voters (agents) can hold only one of two opinions: $\{-1, +1\}$. At each time step:

1. A set of $q$ nodes is chosen at random
2. If they agree a random neighbor copies their opinion; if they disagree the agent flips its opinion with probability $\epsilon_q$

This model reduces to the classical *Voter model* [110] when $q = 1$ and $\epsilon_q = 0$. In this case, we are assuming linear dependence of the opinion change rates of the relative fraction of neighbors in the opposite state; the linear dependence resembles a mechanism of blind imitation, where agents "are infected" by the status of a randomly selected neighbor.) In the q-voter model, consensus in a group (of size q) is necessary for an individual to imitate their neighbors' opinions. This implies a nonlinear dependence on the relative fraction of neighbors in the opposite state and has been used to model competition of species or languages, besides more complex processes of opinion dynamics. This model reduces to the *Sznajd model* [215] when $q = 2$ and $\epsilon_q = 0$. Both the *Sznajd model* and the *Q-Voter model* employ the theory of social impact [141], which takes into account the fact that a group of individuals with the

same opinion can influence their neighbors more than one single individual, similarly to threshold models in spreading processes. The *Sznajd model* also adds a form of information noise $p_{Sz}$ which models the social temperature: the individual has a probability $p_{Sz}$ of making a random decision instead of following the social impact rule (peer pressure, herding,...). When considering only the basic interaction rule on a complete network, all these models reach consensus in the steady state. Even in their simplicity, these models are able to capture some real-world phenomena and are still employed as a baseline for extensions and more complex dynamics. Despite the similarities of these models with spreading, a key difference is that opinion diffusion is up-down symmetric, i.e., the dynamical equations are invariant to the state exchange (x = 0 $\leftrightarrow$ 1). In simple or complex contagion (threshold models) the biological process of infection and recovery or the adoption/resistance to a new innovation are typically different (cf Section 2.2).

### 3.2.2 Majority rule models

A population of $N$ agents takes discrete opinions $\{-1, +1\}$. At each time step:

1. A group of $r$ agents is chosen at random ($r$ can be fixed or drawn from a specific distribution)

2. All agents take the majority opinion within the group (there is a bias towards $+1$ (the *status quo*) in case of tied situations [1]).

This is the basic principle of the *Majority Rule* (MR) model, which was proposed to describe public debates [83] and, in particular, the situation where an initial hostile minority is able to convince the majority [2] favored by a basic and natural mechanism inherent to free public debate. In the context of complex contagion models (cf. Section 2.2), a step function transition rate suggests that individuals will only adopt the "prevailing" opinion within their local network. This concept is akin to the *Majority Rule* model, where agents require a specific count or proportion of their neighbors to hold a contrary view before they alter their own stance.

A common approach in modeling involves treating the opinion variables as real numbers (or vectors with multiple real components). This method aims to depict people's views on various subjects not as a fixed range of options, but as positions on a continuum, with a concept of separation between them. **Models of continuous opinion** may employ an (in-)finite range for their state variables or even periodic boundary conditions where the interval's extremes actually signify the same viewpoint.

### 3.2.3 DeGrootian models.

Suppose a population of $N$ individuals needs to decide the value attributed to a parameter $\theta$ by acting together as a team or committee (i.e., the price of an object) and suppose each individual has his own subjective opinion on the value of the parameter (which can be a point estimate or a probability distribution): in this case, binary opinions are not well suited.

[1] This idea is inspired by the concept of social inertia. In psychology and sociology, social inertia is the resistance to change or the endurance of stable relationships in societies or social groups. It is the opposite of social change. https://bit.ly/3hGBXVz [80]

[2] As it happened with the referendum for the Maastricht agreement in France in 1992 [76] https://bit.ly/3EqhmhW

The *DeGroot* model [55] is a simple opinion dynamics model where opinions are *continuous* in the range $x_i \in [0, 1]$ and the updates are *synchronous*: at each time step individuals do not gain new insights or obtain new information, but can discuss with each other to update their opinion. Therefore, the dynamics is given by:

$$\mathbf{x}(t) = \mathbf{W}x(t-1) = \mathbf{W}^2x(t-2) = \cdots = \mathbf{W}^t x(0) \tag{3.1}$$

where **W** is a row stochastic matrix of the weights agents put on the opinion of their neighbors, which can be interpreted as how much they trust their neighbors' opinions on that subject and, therefore, how much that specific agent will influence their opinion change throughout the dynamic.

The reaching of a consensus is determined by the presence of highly influential agents, i.e., agents who other highly influential individuals trust.

The *Friedkin-Johnsen* model [77, 78] is the first extension of an opinion dynamics model which included the idea of **stubborn agents** that is encoded giving each agent a level of *susceptibility to influence* in $[0, 1]$

$$\mathbf{x}(t+1) = \mathbf{DWx(t)} + \mathbf{(I\text{-}D)x(0)} \tag{3.2}$$

where **D** is the matrix of susceptibility, with susceptibility being (1-$d_i$). Stubborn agents in the *Friedkin-Johnsen* model are those who resist influence from their peers, maintaining their initial opinions throughout the dynamics. Their presence can significantly influence the dynamics of opinion formation, potentially slowing down the convergence to consensus, making the network more resilient to noise, and strategically influencing the outcome of discussions. If the matrix of susceptibility is the identity matrix, the model reduces to the DeGroot model.

### 3.2.4 Bounded confidence models.

While consensus is an interesting phenomenon to understand, it is not the only state worth attention. One of the most famous and earliest ways of preventing trivial consensus is the introduction of **bounded confidence** in opinion dynamics models.

The so-called *Bounded Confidence* models constitute a broad family of models where agents are influenced only by peers having an opinion sufficiently close to theirs, namely below a certain confidence threshold $\epsilon$. This characteristic is justified by sociological theories such as **homophily**. This is the tendency of individuals to associate and bond with similar others in terms of various attributes, such as demographics, beliefs, values, and interests, often summarized with the famous proverb "birds of a feather flock together" [157]. The presence of homophily has been proven in various contexts [161, 82], online social networks being one of the most disputed in recent years – especially when considering the dimension of political opinions [43]. In this context, in fact, homophily is accused of fostering divides and Echo Chambers [43], where people of similar ideologies only interact with each other, reinforcing their opinions and eventually increasing their Polarization. An extended overview on the role of such "bias" and other drivers of such unintended and undesired phenomena is present in Section 4.2.

Bounded confidence, however, can also be interpreted, for example, a lack of understanding, conflicts of interest, or social pressure [54]. This threshold has also been referred to as open-mindedness [150], while some may argue that it is more similar to influenciability. While there are certainly differences between bounded confidence and open-mindedness, for the purposes of this thesis, it may be useful to

consider them as related concepts, and in the remainder of this thesis, we often use the two terms as synonyms.

Two of the milestones in BCMs are the models by Deffuant-Weisbuch [54], and Hegselmann-Krause [104]. Both of these models are grounded in the concept of repeated averaging within the constraints of bounded confidence, but they diverge in their respective communication protocols.

Due to its importance in the present thesis, let's start with the definition of the *Deffuant-Weisbuch model* (*DW Model* model) in its most general form [149].

**Definition 11 (*DW Model*)** *Let there be a population of N agents and an appropriate opinion space $X \in \mathbb{R}^d$. Given an initial profile $\boldsymbol{x(0)} \in \boldsymbol{X}^N$, bounds of confidence $\epsilon_1, \epsilon_2, ..., \epsilon_n > 0$, influence parameters $M = (\mu_{ij})i, j = 1^N$ and a norm $\|.\|$, the* DW Model *is a random process $(x(t))_{t\in\mathbb{N}}$ where at each discrete time step t a pair $(i, j)$ of agents is randomly selected from the population. The selected pair performs the action*

$$x_i(t+1) = \begin{cases} x_i(t) + \mu_{ij}(x_j(t) - x_i(t)) \text{ iff } \left\|x_i(t) - x_j(t)\right\| < \epsilon_i \\ x_i(t) \end{cases}$$

*iff $\left\|x_i(t) - x_j(t)\right\| < \epsilon_i$ to update the opinion of agent i. The same for agent j. If $\epsilon_1 = \epsilon_2 = ... = \epsilon_N$ or if $\mu_{ij} = \mu \forall_{i,j\in\{1,2,...,N\}}$ we call the model homogeneous respectively in bounds of confidence or in influence parameters. Otherwise heterogeneous.*

The parameter $\mu$ represents the influence strength between two agents in the population. Specifically, $\mu_{ij}$ represents the influence that agent $j$ has on agent $i$ when they interact where $\mu_{ij} = 0$ means that agent $i$ trusts so little agent $j$ that they will not change their opinion even if $j$'s falls within their bound of confidence; if $\mu_{ij} = 0.5$ agent $i$ average their opinion with agent $j$'s, however, if the parameter is symmetric, it means instantaneous agreement between the two agents (both taking their average opinion); finally it is normally not studied the presence of influence ratios above 0.5 meaning that both agents value more the other's opinion than their own and – at the extreme case of $\mu_{ij} = 1$ agent i takes agent j's opinion and vice versa.

In this thesis work, we will mainly refer to Deffuant's model in its initial formulation [54]. In that model, the space of opinions is $X \in [0, 1]$, and the parameters are homogeneous over the whole population with $\epsilon \in [0, 1]$ and $\mu \in [0, 0.5]$.

In the homogeneous model with a common confidence bound ($\epsilon$), it has been demonstrated that the system converges to a limiting opinion profile. In this scenario, any pair of opinions $x_i$ and $x_j$ are either equal ($x_i = x_j$) or their difference exceeds the confidence bound ($|x_i - x_j| > \epsilon$), rendering further changes in opinions impossible regardless of the choice of $i$ and $j$. Moreover, it is worth noting that the average opinion across all agents remains conserved throughout the dynamics, but this conservation property is exclusive to the homogeneous case.

**Heterogeneous *DW Model*.** Heterogeneity has been studied as well. The first study taking into account heterogeneity [150] divided the population into two classes of agents, namely the open-minded and the close-minded ones. They studied a set of 192 closed-minded agents ($\epsilon 1 = 0.2$) and eight open-minded agents ($\epsilon 2 = 0.4$) and saw that in the short run, the cluster pattern of the closed-minded (two big clusters) dominated while in the long run, the one of the open-minded (consensus) evolved. Consensus can be achieved by mixing closed- and open-minded agents even if both bounds of confidence are far below the critical value of the consensus transition (e.g., $\epsilon 1 = 0.11$ and $\epsilon 2 = 0.22$). However, a notable dynamic was the drift of open-minded clusters towards those of close-minded agents, which

could amplify any initial asymmetries in the opinion profile. Consequently, the final consensus could significantly deviate from the initial average opinion. Chen et al. [38] considered a version of the *DW Model* where each agent had its own confidence bound. As for the homogeneous model, it is shown that when the *DW Model* is heterogeneous in confidence levels, the network will reach a final state in which any pair of agents either are in agreement or the distance between them is greater than the confidence radius of the two (on complete graphs). Consensus is (almost surely) reached iff $\exists_i, \epsilon_i \geq 1$ for any initial profile $\mathbf{x(0)}$. Heterogeneity can be further exacerbated by placing bounds of confidence on edges instead of nodes. In the model by Shang [199], each edge $(i, j)$ has a level of bounded confidence $\epsilon_{ij}$ representing how agent $i$ is open to agent $j$'s opinion.

The *Hegselmann-Krause model* (HK) [104] also incorporates the concept of bounded confidence but opinion updates are synchronous. In fact, at each time step, a random agent $i$ is selected (like in the *DW Model*), and the set of neighbors $N_{i,\epsilon_{HK}}$ within the confidence bound $\epsilon_{HK}$ is identified. The opinion of the agent, which is a continuous value in $[-1, 1]$, is then updated according to

$$x_i(t+1) = \sum_{j \in N_{i,\epsilon_{HK}}} \frac{x_j(t)}{|N_{i,\epsilon_{HK}}|}; \tag{3.3}$$

the idea is that the opinion of an agent is given by the average opinion of their selected within-bound neighbors.

## 3.3 Opinions with Structures: Bringing Opinion Dynamics Modeling a Step Towards Reality

Complexity is one of the principles governing our world, and the collective behavior of a system can hardly be understood and predicted by considering individual units in isolation. To represent the underlying structure of the population, the first attempts at opinion dynamics modeling employed mainly complete networks, square lattices, or – at most – random graphs with a given density. While the complete network choice can be suitable to some contexts, such as a small group of people discussing a topic, the assumption that every individual is connected to every other individual – or at least can communicate to – is pretty strong, and in real settings, this rarely happens. Researchers have demonstrated that individuals are interconnected through social networks – offline and online – which limit the set of possible interactions and make them subject to network effects, making it necessary to describe such systems through the lens of network theory (see section 2.1.1). In recent years, it has been argued that networks themselves – even temporal, multilayer [3] [132, 23], or feature-rich [4] [121, 45] – cannot be enough to provide a complete description of the system. For example, social mechanisms such as peer pressure cannot be modeled with pairwise interactions because they only arise when a group is

[3] Multilayer networks also consist of nodes and edges, but the nodes exist in separate *layers*, representing different forms of interactions, which connect to form an aspect. Aspects, or stacks of layers, can be used to represent different types of contacts, spatial locations, subsystems, or points in time. The edges between nodes in the same layer of an aspect are called intralayer connections, whereas edges between nodes in different layers are interlayer connections.

[4] Feature-rich networks are defined as complex network models that expose one or more features in addition to the network topology, thus including multilayer and temporal networks, but also attributed networks, location-aware networks, probabilistic networks, and heterogeneous information networks.

considered. Finally, even the assumption of a static network structure may be too stringent. Links are broken and created for reasons that may be (in)dependent on the dynamics taking place on the network. Moreover, pairwise contacts happen at a point in time, and interaction timing distribution is mostly heavy-tailed [13], which is not captured by classical agent-based simulations of opinion dynamics models where timing is mainly Poissonian.

### 3.3.1 Scale-freeness, small-worldness and other topological characteristics playing a role on opinion dynamics

Although the mean-field approximation and regular lattice structures are still commonly used as a starting point, more recent analyses and extensions of established models often utilize generative network models (refer to section 2.1.1) as a foundational structure. In actual social networks, individuals are typically farther apart than in complete networks but closer together than in regular lattices (known as small-worldness) and exhibit diverse connectivity patterns (referred to as scale-freeness). As a result, researchers have begun to investigate the impact of network structure on opinion dynamics. An example of this is the Deffuant-Weisbuch model, which has been analyzed on various network structures, including scale-free networks [228], where the threshold of consensus, i.e., $\epsilon > 0.5$ still holds [73]. Building on these findings, researchers have explored the interplay of scale-freeness with other realistic elements, such as the presence of directed edges [210], heterogeneous influence rates [100], discretized opinions [211] (including in the multidimensional case) [124, 123, 3], different psychological characteristics of agents [139], and stubbornness [117]. While the baseline model produces Polarization primarily as a result of "close-mindedness," taking into account various factors such as scale-freeness, specific initial distributions, other psychological factors, and the presence of opinion leaders [128] or informed agents [1] can lead to different outcomes.

While scale-freeness has significant implications for the diffusion of influence, other network characteristics found in the real world are also crucial to consider when studying how opinions form and evolve. For example, Meng et al. [158] investigated the Deffuant model on various network structures, such as deterministic synthetic networks, random synthetic networks, and social networks constructed from Facebook data. Similarly, Gandica et al. [85] applied the same model to small-world networks, specifically Watts-Strogatz (un)directed networks [226], as well as Amblard and Deffuant [5].

### 3.3.2 Relationships are not static: opinion dynamics on evolving and co-evolving topologies

In recent years, scientists have increasingly utilized dynamic networks (cf. Section 2.1.3) as a modeling framework for studying spreading phenomena, such as opinion diffusion. While incorporating realistic contact patterns, previously mentioned studies miss the key ingredient that people's connections change over time due to various reasons; just think of the formation and dissolution of friendships, bounded time spent with schoolmates, and following and unfollowing users on online social media.

Temporal networks, which account for the sequence of interactions, have recently emerged as an effective model for human interactions due to their time-varying "proximity". These networks have shown that temporal patterns such as contact orders, burstiness, and heterogeneous lifetimes can significantly influence

dynamics, leading to behaviors that differ greatly from those observed in static network representations. In binary models of opinion dynamics – as in similar models of epidemics – wider inter-event time distributions slow down the ordering dynamics [67, 216]; moreover, studies on continuous opinions suggested that aggregated networks consistently overestimate the speed of consensus formation [148] with respect to empirical contact sequences. From [42], it also emerges that a higher influence on the final state of the population (average opinion at steady state) is exerted by agents with a longer waiting inter-interaction time.

The principle of homophily plays a crucial role in the coevolution of opinions and connections. While we can only be influenced by connections that exist, it is also true that some of these connections are formed or broken specifically due to changes in our belief systems and opinions. This is because individuals tend to seek out and maintain relationships with others who share similar beliefs, values, and interests. As a result, the process of opinion and interaction/relationship evolution are inevitably tied in the long run. To include such interplay in diffusion models, the concept of adaptive or coevolving networks has been introduced (cf. Section 2.2.3), and a vast literature on adaptive versions of classical models has been rapidly evolving in recent years.

Due to the additional complexity of studying such interplay, starting efforts focused on the Voter Model (cf. Section 3.2.1). In the adaptive version, pairs of voters can either change their opinions or break their connections [233, 92]. The concept of "rewiring" a link – when a connection is not "satisfying" – was introduced by Nardini et al. [167] by connecting agents in a random network instead of a complete one. The concept of rewiring due to discording interactions will be encountered again in Chapter 7.

Bounded confidence models (e.g., *DW Model*) somehow already incorporate the idea of homophilic behavior in the sense that agents only interact with others whose opinions are close to their own. One might wonder why an agent would stay connected with someone with whom he or she disagrees when, for example, environments such as social networks allow us to connect with people similar to us and break them with people we think are too different.

A simple way to model this is to introduce a rewiring mechanism into the model, where agents can break their connections and form new ones – possibly with others whose opinions are closer to their own [130] – [135]. A random agent $i$ – with a certain probability $p$ – will rewire the link with agent $j$ to another agent $z$ if $i$ and $j$'s opinion distance exceed the confidence bound (with probability $1 - p$ a normal *DW Model* interaction happens), or a different tolerance threshold [130].

Sequential processes of rewiring and social influence are studied in [224], similarly to [86]. In [86] first the network grows according to both preferential attachment and and homophily, then a bounded confidence opinion evolution unfolds on the network.

The model predicts the same phenomena in [194], showing how the human tendency to be influenced by information and opinions to which one is exposed and the dislike of disagreeable social ties facilitated by social media may lead to the formation of Echo Chambers. Social influence and rewiring can create completely segregated and polarized Echo Chambers, and this phenomenon is accelerated in the presence of both strong influence and common unfollowing.

### 3.3.3 Peer pressure and other higher-order effects on opinion dynamics

The importance of higher-order structures in social systems has been highlighted in Section 2.1.2.

In the last decade, there has been a growing interest in the study of opinion dynamics on higher-order structures, such as simplicial complexes [113, 195, 193] and hypergraphs [105, 168].

The idea that a "group" of "concording" agents is necessary to influence a single agent has already been developed in the framework of threshold models (cf. Section 2.2). Moreover, group interactions have already been considered in some opinion dynamics models family, i.e., the majority rule models [83], but the underlying structure employed is still a network.

In [113], the idea of a dynamic (adaptive) topology is combined with higher-order structures (simplicial complexes) to formulate an adaptive voter model where the nodes are also subject to peer pressure if they are part of the same simplex.

The model starts from the classical adaptive voter model in which, in addition to "persuasion" events, "reconnection" events are also considered in which links between nodes with opposite opinions are redirected to nodes with the same opinion (again, in fact, opinions can take values exclusively in the set $\{-1, +1\}$). However, in the present model, the authors consider higher-order interactions through which the social mechanism of peer pressure can be modeled. According to this mechanism, if three individuals are connected by a bond of friendship and a situation of disagreement arises in the group, the majority opinion within the group will likely prevail. The model proposes a minimal extension of the adaptive voter model in which to consider a triangle of agents (i.e., a subgraph of 3 vertices fully connected) a group of friends, using an additional structure beyond nodes and vertices. The chosen structure is the 2-simplex: a triangle of nodes forms a friendship if there is a two-dimensional symplex between them beyond the one-dimensional symplexes (the arcs). Since nodes can have an opinion only in the set $-1, +1$, within that structure, there is always a well-defined majority towards one or the other opinion.

At each time step, an arc is chosen randomly:

1. If the arc is not part of a two-dimensional simplex, the rule of the classical adaptive voter model applies, i.e., if the nodes have opposite opinions, then with probability $p \in [0, 1]$ one of the vertices (chosen with equal probability) reconnects the arc to a node with its own opinion, chosen randomly from the remaining vertices, otherwise with probability $1 - p$ one of the two randomly chosen vertices changes its opinion and adapts it to the opinion of the other; if the two nodes have the same opinion, nothing happens.

2. If the arc is part of at least one two-dimensional simplex and the two nodes have opposite opinions, then with probability $q \in [0, 1]$, a simplex is randomly chosen, and the majority persuades the minority with probability $p$. With probability $q - 1$, the classical adaptive voter model is applied instead, and then the arcs are reconnected.

3. Nothing happens if the arc is part of at least one two-dimensional simplex but is inactive.

Regarding the creation of a simplicial complex from a network $(V, E)$, in the present model, the authors chose to take the initial population of triangles in the graph, and a set $S$ of these are declared 2-simplexes while the remaining are simple

triangles. Since some of the simplexes might be destroyed through a rewiring event, each time this happens, another triangle is chosen to become a two-dimensional simplex. The simulations performed on such a model show that the heuristic that peer pressure leads to Polarization is true.

The dynamic rule of the Deffuant-Weisbuch model for pairwise interactions (Definition 11) states that at each time step, the two agents' opinion is updated only if their opinion difference is less than the confidence bound.

If we change the underlying social structure of the model from a graph to a hypergraph, interacting pairs become interacting groups modeled by hyperedges.

In the HOID model [195], the dynamic rule determines that, at each time step, a random hyperedge $h$ is selected, and every node included in $h$ updates its opinion with the following rule:

$$x_i(t+1) = \begin{cases} \overline{x_h} & \text{if } max_{j\in h} x_j(t) - min_{j\in h} x_j(t) \leq \epsilon \\ x_i(t) & \text{otherwise} \end{cases} \tag{3.4}$$

Where $\overline{x_h} = \frac{1}{|h|}\sum_{j\in h} x_j$ is the average opinion of the agents linked by the hyperedge $h$. This means that the opinion update happens only if the opinion differences between all the interacting neighbors are less than the given confidence bound $\epsilon$. In other words, a context influences the agents' opinions only if it does not include users with opinions too distant with respect to the rest of the group, which, instead, precludes the possibility of reaching a consensus. This kind of group interaction is different from the one proposed in the Hegselman-Krause model: the difference stands in the fact that in this higher-order model, the distances between all nodes within the hyperedge matter, while in the HK model, only the distance between a given node and its neighbor have an impact on the opinion change of the target node. Consequently, a dissenter can block the interaction of all other agents in the hyperedge in the HOID model, a mechanism absent from the HK model.

Like with pairwise interactions, the confidence bound value is fundamental and has a great impact on the outcome of the model implementation. A small confidence bound prevents many group discussions from being influential, especially the more confrontational ones. Moreover, with low confidence bound, the probability of an interaction to be influential decays exponentially with respect to the size $|h|$ of the hyperedge since larger groups of agents with very different opinions have a lower probability of reaching a consensus. While higher confidence bound levels bring consensus, low values cause opinion fragmentation and Polarization. The interesting insight drawn from [195] is that with the hypergraph configuration, there is not a sharp transition to a consensus like in [54].

Similarly, in [105], a Hypergraph Bounded Confidence Model (HBCM) is defined. In this case, to generalize the notion of confidence bound to hyperedges, authors define a *discordance function* that maps a hyperedge and an opinion state to a real number and quantifies the level of disagreement among the nodes that are incident to a hyperedge, to determine whether or not these nodes update their opinions. Authors in [105] consider the following family of discordance functions:

$$d_\alpha(h, x) = (\frac{1}{|h|-1})^\alpha \sum_{i\in h}(x_i - \overline{x}_h)^2 \tag{3.5}$$

which is parameterized by the scalar $\alpha$, where $\overline{x}_h = \sum_{i\in h} x_i/|h|$. If the discordance $d_\alpha(h, x(t))$ is less than the confidence bound $\epsilon$, the hyperedge $h$ e is concordant at time $t$. Otherwise, it is discordant. In the case of $\alpha = 1$, $d_1(h, x(t))$ is equal to the

unbiased sample variance of the opinions of the nodes that are incident to $h$. The scaling parameter $\frac{1}{|h|-1}$ prevents advantaging hyperedges with few nodes over ones with many nodes when there is an opinion update. In the model, at each time step, a hyperedge $h \in H$ is selected at random according to some probability distribution (e.g., uniform distribution). If the discordance function is less than the confidence bound $\epsilon$ th nodes $i \in h$ update their opinions $x_i$ to the mean opinion $\overline{x}_h$; otherwise, their opinions do not change. One way to think about this update is that nodes are "peer pressured" into conforming to the mean opinion of the group when the overall discordance of the group is sufficiently small.

More formally:

$$x_i(t+1) = \begin{cases} \overline{x_h} & \text{if } i \in h \text{ and } d(h,x) \leq \epsilon \\ x_i(t) & \text{otherwise} \end{cases} \tag{3.6}$$

## 3.4 Opinions with External Information

The first generation of opinion formation models lacked a crucial element in the process – the presence of mass media or an information environment. As pointed out by several studies [156, 33, 217, 188, 114, 28, 94, 189], the impact of mass media on shaping our culture, ideology, and opinion spectrum cannot be ignored. While social influence (i.e., word of mouth) plays a significant role in opinion formation, news, or political propaganda, exposure can also affect our belief system, e.g., convincing an initially skeptic individual to wear masks and practice social distancing during a pandemic [4].

Generally speaking, most models incorporating mass media add a probability of interacting with such external agent $p_m$ or adopting its opinion.

Different dynamical regimes – such as fragmentation or disorder – emerge by adding an additional layer of mass media communication [156] in the *DW Model*, also when considering additional repulsive behaviors [154] or heterogeneous bounds of confidence [188].

The most straightforward extension of the *DW Model* comes from [33]; here, the whole population interacts with a single mass media every T generations, showing that an open-minded population converges to the external opinion. If populations are close-minded, the final state exhibits richer dynamics based on $\epsilon$, $T$, and the value of the promoted opinion. To ensure pluralism and prevent conformity, a plurality of media with different orientations is fundamental [25].

In [188], the authors analyzed two cases: one with only two bounds of confidence and another where each individual has their own characteristic level of confidence. The interaction of individuals with the mass media is determined by their confidence level and the mass media intensity ($p_m$). With probability $p_m$, if the difference between an individual's opinion and the mass media's opinion is less than the individual's confidence level, the individual interacts with the mass media and updates their opinion accordingly. With probability $1 - p_m$, the individual interacts with others in the system based on their confidence levels and updates their opinion. The mass media's persuasion capacity is analyzed by calculating how often it persuades more than half of the population to follow its opinion. The study shows that the persuasion capacity of the mass media is optimal for intermediate levels of heterogeneity and is sensitive to the initial conditions and parameter values. The authors also found a counter-intuitive effect where the persuasion capacity of the

mass media decreases if its intensity is too large. In Chapter 5, we will see that the implementation of the *ABMM Model* resembles this model.

Research on German political propaganda [24] and the impact of media on Brexit [227] highlights the importance of considering external influence as a force capable of steering the opinion evolution towards a desired goal. In addition to the natural factors that influence opinion formation, external agents such as governments, companies, or terrorist groups may have a vested interest in shaping public opinion on a particular topic or product [53], e.g., influencing the adoption of innovations. These agents may use propaganda or other tactics to promote a specific opinion over others [74] or [151] to achieve a certain value for the consensus opinion through their actions. Moreover, external agents may also try to prevent people from reaching more extreme opinions to mitigate potential risks [217]. This could be achieved through various means, such as limiting access to certain information or using persuasive techniques to steer public opinion towards a more moderate stance. The optimal behavior – however – may be counter-intuitive: e.g., an aggressive media campaign might fragment the population, and the desired goal would become impossible to reach [87]. The potential risk of such models has been highlighted in [209], "serving and guiding commercial companies, politicians, populist movements (such as the anti-vaccination activists), etc.".

Besides their potential to manipulate public opinion, mass media are argued to be one of the drivers of the rising Polarization in Western societies (polarizing effects of news media (e.g., McLaughlin,2018)), particularly through the enhancement of Echo Chambers [57]. The impact of media on political Polarization is complex and not always straightforward. While some studies suggest that media fragmentation and partisanship contribute to increased ideological and affective Polarization among people [142], others argue that media may not always have a significant effect on Polarization [221, 219]. Additionally, there is evidence to suggest that under certain conditions, exposure to political information can actually reduce Polarization [136, 230].

Overall, the mixed findings indicate a need for further examination and evaluation of the existing research on this topic.

While there have been prominent discussions around (political) Polarization in the last 20 years, it needs to be acknowledged that in the last few years, we saw the rise of Polarization [5] – with the related Echo Chamber phenomena – increased social media use, more partisan media, and other elements that together contribute to the creation of a biased or Polluted Information Environment. We will dig deeper in Chapter 4 into the main characteristics of such biased environments and their possible effects on the process of opinion evolution, which is one of the key research questions behind this thesis.

[5] Pew Research Center - U.S. Politics & Policy (2017). *The Partisan Divide on Political Values Grows Even Wider*. Geiger, Abigail

# Chapter 4

# Biased Societies: The Role Of Biases In Polluting Information Systems

In contemporary society, the manner in which individuals access and consume information has undergone a significant transformation. Over the past decade, social media platforms have become an indispensable aspect of daily life. These platforms facilitate the sharing of information, expression of opinions, and interaction with peers. Consequently, the ways in which individuals interact, consume information, and communicate with others have been fundamentally altered. Traditional media, such as newspapers and television programs, rely on human editors to curate content for large consumer groups. As a result, there are numerous consumers but a relatively limited number of newspapers and television programs. In contrast, social media platforms employ recommender systems that function as editors, selecting content tailored to each user's individual preferences. This personalized approach has significantly impacted the dissemination of information within society, marking a departure from the methods employed by traditional media outlets.

Digital age communication has evolved into new paradigms characterized by multiplicity, interactivity, the absence of space-time barriers, and, most notably, an unprecedented level of freedom of expression. When users post content online, they often move beyond one-to-one communication, fostering an interactive exchange of opinions with a potentially infinite user base. Sociologists attribute the success of social media platforms to these blurred boundaries, which result in interactions that are immediate, global, and heterogeneous. Every user has the opportunity to express their opinion on various debates or disseminate news, sometimes even becoming a citizen journalist or going viral for a day. Individuals now have the power to shape public discourse and influence opinions on a global scale. This shift has led to a more decentralized flow of information, where traditional gatekeepers, such as news organizations and editors, no longer have exclusive control over the narrative. Consequently, the influence of social media on public opinion has become a critical area of study for researchers, policymakers, and media professionals alike.

While the freedom of expression and lack of boundaries on social media platforms have led to numerous benefits, they also come with significant drawbacks. In the early days, it was argued that the advent of the Internet, guaranteeing free access to a huge amount of information, would be a boon for democracy. It is not disputed that social media have some advantages, in this sense: information is freed from the barriers of classical journalism and mainstream media, every user can be a "journalist" and promote their points of view, multiplying information sources and contents. However, as information becomes more accessible, it also becomes more

open to influences from non-traditional actors – in most contexts, anyone can create and disseminate information. Online platforms can inadvertently increase the likelihood of encountering malicious behaviors or content and (un)intentionally enhance harmful phenomena like Polarization and Echo Chambers, creating what is called a **Polluted Information Environment**.

**Definition 12 (Pollution)** *In the context of online social media, we define pollution as "the (measurable) effects of an endogenous/exogenous phenomenon that deviates/tempers the unfolding of a public debate/opinion formation process". Some examples of widely known **polluting phenomena** are d/misinformation campaigns, coordinated misbehavior (e.g., a political botnet), cyberbullying, biased content/interaction recommendations, exposure to extremist agents/media, etc.*

**Definition 13 (Polluted Information Environments – PIEs)** *We define PIEs as those online environments where the existence of polluted contents and behaviors biases opinions, exchanges, and public debates.*

**Pollution effects: emergence of Echo Chambers, ideological Polarization, and Filter Bubbles** Before delving into the description of some of the possible drivers of information pollution in online social networks, let us stop to focus on some definitions of the main concerns related to these realities.

**Definition 14 (Filter Bubble)** *The term Filter Bubble traditionally denotes the sphere of information that an online user can access, which is argued to be steered and excessively personalized by algorithms governing digital platforms.*

Eli Pariser first introduced the concept of Filter Bubbles, suggesting that recommendation algorithms limit the diversity of content users encounter, creating a personalized sphere of interests and search preferences [181]. The theory posits that continuous exposure to content that aligns with one's existing beliefs can lead to further *Polarization* and association with like-minded individuals. However, this conclusion hinges on the assumptions made about the cognitive processes involved in opinion change (e.g., motivated reasoning and selective exposure).

In general, the term **Polarization** refers to the division into two sharply contrasting groups – or any number of politically relevant dimensions – or sets of opinions or beliefs in other domains [155].

This term is usually used in the context of the social or political domain, where it indicates the divergence of political or social attitudes towards ideological extremes, ultimately leading to "partisan" attachments; as such is driven by a range of factors from socio-political issues, to cultural, social, and economic changes.

The political situation in the United States in recent decades serves as a clear illustration of this phenomenon. Indeed, the survey conducted by the Pew Research Center [36] unequivocally indicates a steady escalation of the ideological division between the political parties. Particular concerns, in this domain, arise over *affective* form of Polarization. The concerns rising on this increasing Polarization are due to the idea that a certain extent of "social cohesion" is necessary for democracies to function, e.g., for governments to perform actionable changes.

The relationship between online platforms and Polarization is complex and unresolved. Recent research suggests media usage amplifies Polarization, but methodological issues exist, such as non-operational definitions and studies focusing on single platforms (e.g., Twitter) [136]. Other studies suggest the implication direction

is reversed, i.e., polarization levels influence social media use [173], and platforms may enhance divisive content, leading to misconceptions about out-groups [10].

When polarized views on controversial issues are shared and reinforced among a group of individuals, **Echo Chambers** (as defined by Sunstein [214]) emerge. Broadly speaking, the term Echo Chamber refers to a situation in which beliefs are amplified or reinforced through repeated communication within a closed system, insulated from opposing perspectives. Thus, the definition of this phenomena [43] is mainly based on the theory of selective exposure [133] (with selection operated both at user-level and at platform-level through filtering algorithms) and of confirmation bias [170]. However, the concerns on Echo Chambers and Filter Bubbles go further, suggesting that this homophilic selection will not only confirm pre-existing beliefs but will reinforce people's views, exacerbating extremis.

Despite the prevalence of the phenomenon, a formal definition of Echo Chambers remains elusive. It is important to note that both Polarization and Echo Chambers (and one may argue also Filter Bubbles, to some extent) exist in real-world situations as well and have been discussed in the socio-political literature for decades; the discourse on Echo Chambers, for example, is clearly tied to concerns about fragmentation of public discourse, existing well before the digital age.

However, it is also worth noting that Polarization and Echo Chambers typically arise around controversial issues, such as societal or political topics, and are not harmful phenomena per se. The internet is replete with online groups of people united by shared interests, such as a specific genre of music, a sports team, or even homophilic groups formed around individuals facing particular challenges, like mental health issues.

Before delving into the technical aspects of these phenomena, it is prudent to address a recent surge of skepticism among researchers [62, 30]. The topic of discussion is characterized by an absence of comprehensive definitions for both the Filter Bubble and Echo Chamber phenomena, as well as of universally applicable methodologies for assessing their existence.

However, despite differing viewpoints, we cannot ignore that the influence of social media on our daily lives has concerning implications that call for further scrutiny.

To better understand the motivations behind the emergence of such polluted realities, it is essential to consider to what extent such phenomena are solely a consequence of social media or if they reflect socio-psychological aspects and have always existed, albeit on a smaller scale.

In the contemporary media landscape, consumers are presented with an overwhelming array of information across various formats. This necessitates the need for consumers to make discerning choices about their information intake.

**Definition 15 (Information Proliferation)** *Information proliferation refers to the rapid growth and dissemination of information, both structured and unstructured, in contemporary society.*

This phenomenon is driven by advancements in technology, such as social media, mobile devices, and the internet, which enable individuals to access, create, and share information more easily and quickly than ever before [106].

This unprecedented flow of information overwhelms users (people); they cannot digest this information diet in its integrity. They are forced to **select** what to process and eventually share. This selection process is inevitably influenced by a number of

biases, e.g., cognitive and algorithmic, which – as we suggested in the brief discussion above – are argued to be among the main drivers of these harmful phenomena (Polarization, Echo Chamber, etc.).

**Cognitive biases** It is obvious that we have a limited amount of time and attention [108] to dedicate to information (whether it is gathering information, reasoning, or discussing it with our peers): this means that humans or – in the online world – users are exposed to a limited amount of content and interactions. However, the choice of which users/content to engage with is not made with the aim of having a balanced diet, but it is highly affected by cognitive biases[1]. For example, when there is too much information, and we can choose what to focus on, we tend to concentrate on the pieces that confirm our own existing beliefs and ignore details that may contradict our own beliefs [2]. Individuals tend to overcome this psychological distress through selective exposure [102] and confirmation bias [170], meaning they select and disseminate information that reinforces their pre-existing ideologies while avoiding opposing viewpoints. This behavior can be only exacerbated by the information overload experienced daily [106]. An experiment from 2009 found that when presented with different news articles, people tend to select news based on anticipated agreement, i.e., news from sources they know to be closer to their leaning [122]. Evidence of political Polarization and selective exposure was also found on blogs [143] and social networks like Twitter [89].

By understanding the role of individual biases and the psychological mechanisms behind information selection and endorsement, we can better comprehend the factors that contribute to the formation and reinforcement of opinions in both online and offline environments.

These biases can affect our "opinion-making" process, as they may cause us to overlook pivotal information or focus only on evidence that confirms our assumptions. As much of the literature on Echo Chambers suggests, in the context of social media and news consumption, confirmation bias can lead to the reinforcement of pre-existing beliefs and attitudes, making people more susceptible to dis/misinformation and polarized views.

Theoretical models [54] suggest that when confirmation bias is strong enough, it can lead to Polarization even without other contributing factors. When other biases, such as selection bias, interact with confirmation bias, the impact on opinion formation and Polarization can be more significant.

**Birds of a feather flock together** Individual biases can easily translate into group biases when individuals consciously or unconsciously choose with whom to interact and communicate.

One factor that influences the formation of friendships and interactions is homophily, one of the most robustly documented social phenomena.

**Definition 16 (Homophily)** *Homophily is the tendency of individuals to associate and bond with others who are similar to them on some dimensions [157].*

Various dimensions contribute to homophily, such as gender [201], age [206], geographical location, shared habits [72], attitudes towards life [232]. Both homophily

[1] https://bit.ly/3EtGWCT

[2] Cognitive dissonance is the process for which people experience discomfort when presented with information that challenges their beliefs or decisions [71]

and its counterpart heterophily act as fundamental principles in the choice of people's social circles [197, 65]. Such relationships between node interactivity and the properties carried by the nodes can be extremely useful in diffusion processes and opinion contagion dynamics [6]. Academic discourse reveals little doubt that homophily exists as a grouping tendency in humans, but many questions remain as to how homophilic relationships are formed and the effects of these relationships. As emerges from these examples, homophily may act on two levels: there is homophily that is *induced* by structural constraints and opportunities, and there is homophily that is *chosen* according to preferences. In off and online social networks, selective exposure led by cognitive dissonance and confirmation biases [70], in fact, does not only affect the information users choose to interact with but also the composition of their network of interactions. While, in some platforms, this is mainly determined by real-life social ties (family, friends, colleagues...), in other contexts, people tend to create their personal bubble of like-minded individuals – who they may not even know in real life – to create a comfort zone where there is no disagreement nor conflict.

It emerges that homophily – reducing diversity in information and network reach – is an essential ingredient of the fragmentation of (political) discourses and Echo Chambers phenomena [116, 115, 166, 160, 7].

**Media biases** Besides social interaction, the information environment is largely constituted by external sources (mainly in the form of mass media broadcasts), enhancing awareness of e.g., socio-political issues and events [95, 114]. The importance of mass media has been widely recognized, and traditional mass media have been argued to influence individual and public health [29, 225] on issues ranging from eating disorders [218], tobacco consumption [81], and vaccinations [200]. Moreover, news articles, TV news, and political talk shows all play a central role in shaping opinions.

Especially when it comes to the communication of political information – essential for informed electoral decisions [34] –, mass media serve as the central platforms for political discourse and primary source of political information [68], holding the power to manipulate how people think about internal and international politics [207].

Given this power, one would hope that mass media would at least provide their audiences with reliable news and truthful representations. Nonetheless, media coverage often exhibits an internal bias, reflected in the news and commonly referred to as media bias [101]. Various definitions of media bias and its specific forms exist, each depending on the particular context and research questions studied.

**Definition 17** *Media bias (slanted news coverage) is an internal bias in media coverage, both intentional and systematic [229].*

Media bias can manifest in various forms, influencing the way stories are presented and perceived. For instance, an article may exhibit bias by intentionally promoting a specific opinion on a topic or by crafting a memorable narrative, often through the emphasis on particular aspects or the use of sensational language [32]. In the political realm, unbalanced coverage can lead to bias by disproportionately focusing on certain topics or entities, thereby skewing the overall representation [202]. Agenda setting is another prevalent political practice that leverages media outlets to shape public discourse. By selectively reporting or concealing information, media outlets can influence the overarching narrative. According to a study on the impact

of media bias, editorial slant – a measure of the quantity and tone of a media outlet's candidate coverage as influenced by its editorial position – has been found to affect voter behavior [61]. In this context, Brockman and colleagues which investigated the impact of partisan media on voting behavior [27]. The researchers recruited a sample of regular Fox News viewers, a right-wing media outlet, and incentivized them to watch CNN, a left-wing media outlet, for a month. The findings of this study are particularly intriguing, as they reveal substantial effects of watching CNN instead of Fox News on participants' factual perceptions of current events, such as the COVID-19 pandemic and the positions of the 2020 presidential candidates. The two media outlets exhibited markedly different topic coverage. Fox News appeared to downplay negative information about then-President Trump and the severity of the coronavirus, while CNN did not extensively cover protests against racism or criticize then-candidate Biden and the Democrats.

Factors influencing this bias include ownership or a specific political or ideological stance of the outlet and its target audience [3]. Media choices can also be influenced by their profit-oriented nature, leading to content selection aligned with the audience's interests that fuels this profit, disregarding issues and problems (and portions of the population, such as minorities) that would guarantee fewer earnings [63].

Undoubtedly, media bias constitutes a significant factor influencing the quality of any information environment, contributing to further polarizing an already biased population towards more extreme positions to increase their revenues, satisfy their investors' needs, and help politicians set their agendas.

**Algorithmic biases** The characterization of the present information environment – from the perspective of biases – is not yet complete.

Cognitive, relationship, and media biases have been deeply ingrained in human nature, likely since the dawn of human interaction, shaping our social interactions, decision-making processes, and the way we consume and share information. In the digital age, these biases have become even more pronounced, as algorithms and artificial intelligence systems can – arguably – amplify and perpetuate them.

The proliferation of information in the digital age has made it essential for algorithms to filter and recommend content, easing user navigation in the vast sea of online information, which would otherwise be overwhelming. In 2013, Facebook disclosed that an average user's News Feed could display around 1,500 stories, but only 300 are selected based on factors like user interaction, engagement metrics, past behavior, and user actions such as hiding or reporting posts [4].

Recommender systems on online social platforms often aim to maximize user engagement as one of their primary objectives. User engagement refers to the level of interaction and involvement that users have with the platform, which can be measured through various indicators such as time spent on the platform, number of clicks, likes, shares, and comments. By providing personalized and relevant recommendations, recommender systems can increase user satisfaction, encourage users to spend more time on the platform, and foster interactions among users.

It is important to note that focusing solely on maximizing engagement can have some unintended consequences. For instance, recommender systems that prioritize engagement may inadvertently promote content that is controversial, polarizing, or sensational, as such content tends to generate more interactions and reactions from

[3]Reuters Institute (2022). *Digital News Report*.

[4]Meta for Business (2013). *News Feed FYI*. https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed

users. For example, simulations in [37] showed how focusing solely on user engagement can lead to overexposure to negative content, Polarization, and concentration of social power in the hands of the most toxic users.

This can lead – as argued by a number of scholars – to the amplification of Echo Chambers, Filter Bubbles, and the spread of dis/misinformation, which can have negative effects on the overall health of the online social network, contributing to the creation and maintenance of a Polluted Information Environment.

Besides the already mentioned problems, algorithmic recommendations – in conjunction with individual choices – have other unintended (social) consequences, e.g., an enhancement of preferential attachment mechanisms, boosting the popularity of already popular accounts on Twitter [213].

While substantial evidence exists indicating that recommender systems may contribute to the emergence of polarization, echo chambers, and filter bubbles, establishing a direct causal link between these phenomena and recommender systems poses a significant challenge.

Despite these formidable challenges, numerous studies have provided indirect evidence of the relationship between recommender systems and these harmful consequences. These studies often rely on observational data, simulations, or controlled experiments with limited scope. While they may not definitively establish a direct causal link, they strongly suggest that recommender systems play a significant role in shaping the information landscape and can contribute to the emergence of these phenomena.

Despite the lack of understanding of the relationship between recommender systems and harmful consequences at individual and societal levels, possible solutions to tackle the ethical concerns around these technologies are already present in the literature.

In [88], authors study algorithmic techniques for dismantling Echo Chambers, connecting users with opposing views.

## 4.1 Models of pollution

As was to be expected, the last 20 years have seen the proliferation of opinion dynamics models that have tried to incorporate elements of information pollution, recreating synthetic PIEs in a certain sense, or have, in any case, tried to develop models that propose mechanisms that reproduce pollution phenomena such as Echo Chambers and Filter Bubbles, suggesting that they depend on factors intrinsic to information exchange and not distortions introduced by the evolution of the way we interact and communicate.

In the following, we will introduce some of the main opinion dynamics models that attempt to incorporate specific characteristics of online social networks, in particular, the presence of recommender systems and algorithmic bias, which is one of the cornerstones of the work developed in this thesis.

One of the initial efforts at comprehending the impact of biases in opinion evolution is the model proposed by Deffuant and Weisbuch [54]. The bias is introduced through a single parameter, known as the "confidence bound" or "confidence threshold", which shows that agents exclusively influence each other when their opinion distance is less than a specific threshold.

However, in recent years, there has been significant debate surrounding the role of social media platforms in society.

A key question is whether the recommender systems deployed by platforms to deliver content to users have any tendency to foster Polarization/radicalization and phenomena of Echo Chambers and Filter Bubbles. Already in 2015, [152] argued that personalization algorithms would foster Polarization under the assumption that opinions are reinforced when we interact with like-minded individuals, while they would actually weaken Polarization if we assume rejection on discordant interactions, calling for more empirical research on the assumptions.

The real mechanisms are far from being assessed up to this day.

**Algorithmic bias amplifies opinion fragmentation and Polarization: a bounded confidence model** Deffuant-Weisbuch model already predicted Polarization when the confidence threshold was low. In 2019, Sirbu and colleagues [204] studied the effects of biasing interactions towards like-minded individuals in a bounded confidence model.

**Definition 18 (Algorithmic Bias model – *AB Model*)** *Let us assume a population of $N$ agents, where each agent $i$ has a continuous opinion $x_i \in [0, 1]$. At every discrete time step, an agent $i$ is randomly picked from the population, while $j$ is chosen from $i$'s peers according to the following rule:*

$$p_i(j) = \frac{d_{ij}^{-\gamma}}{\sum_{k \neq i} d_{ik}^{-\gamma}} \tag{4.1}$$

*If their opinion distance is lower than a threshold $\epsilon$, $|x_i - x_j| \leq \epsilon$, then both of them change their opinion according to Eq. 11.*

The AB model introduces another parameter to model the algorithmic bias: $\gamma \geq 0$. This parameter represents the filtering power of a generic recommendation algorithm: if it is close to 0, the agent has the same probability of interacting with all its peers. As $\gamma$ grows, so does the probability of interacting with agents holding similar opinions while interacting with those who hold distant opinions decreases. Therefore, this extended model modifies the rule to choose the interacting pair $(i, j)$ to simulate a filtering algorithm's presence.

$d_{ij} = |x_i - x_j|$ is the opinion distance between agents $i$ and $j$, so that for $\gamma = 0.0$ the model goes back to the DW-model, i.e., the interacting peer $j$ is chosen at random from $i's$ neighbors or – in other words – every neighbor is assigned the same probability to be chosen.

Results showed (on a fully connected network) how algorithmic bias, i.e., the tendency to interact more with similar opinions *because of* algorithmic personalization, may induce fragmentation and Polarization even in settings where the baseline model predicted consensus. Also, clusters tend to be more distant, and consensus needs more time to be reached.

The task of understanding the role of recommender systems on opinion evolution and its interplay with various drivers of such dynamics has been tackled with a variety of different approaches, reaching different results.

Primarily, these models support the central assertion of [204], which states that recommender systems contribute to Polarization and diminish content diversity [90]. Polarization escalation has been corroborated in [186, 185, 183] within the realm of binary opinions, involving a blend of "similarity bias" and "popularity bias" [21], where the debate space is marked by ambiguity [60] and influenced by socio-cognitive biases [220]. As previously mentioned in Chapter 4, homophily could be a

crucial factor in the formation of Echo Chambers, an effect that may be intensified by people recommender systems, as confirmed by [44]. Another intersecting aspect is the presence of "influencers" within online social networks (OSNs). In this context, algorithmic personalization may not only promote personalization but also facilitate the development of Echo Chambers surrounding structurally privileged influencers [84].

It is worth noting that, while audits and models show that recommender systems increase the levels of pollution, studies on real data suggest that these are not the primary drivers of Polarization and radicalization; this apparent "paradox" may arise due to models not accounting for users choices, which are rarely to consume niche content (e.g., extremist content) [191]. Authors therefore stress the importance of modeling user choices and call for a nuanced interpretation of "algorithmic amplification". Other studies suggest that – even if cognitive/algorithms did not play a role – the ecosystem's characteristics would be sufficient in creating Echo Chambers [75].

From this discussion, the only certain thing that emerges is that modeling personalization algorithms is not simple because we don't know much about how these algorithms work in real settings yet.

## 4.2 Explaining pollution: bridging the gap between models and data

Although assumptions and simplifications are made in building the presented opinion dynamics models, they have proven very useful in explaining well-known phenomena in opinion formation. This literature on opinion dynamics models is wide, going from binary opinions and pair-wise interaction models to continuous opinions on time-evolving higher-order systems, trying to narrow the gap between the models and the real systems from a theoretical perspective.

Despite online social networks offering such a huge opportunity to retrieve people's opinions, friendships, interactions, and discussions, there is still a lack of quantitative analysis of real data, and empirical approaches are claimed to be the next necessary step by many researchers in the field. Application to real data to validate models' conclusions is still very scarce, and the lack of this kind of approach is one of the two major issues addressed in our work.

**Common data sources** Online Social Networks (OSNs) like Facebook, Reddit, Twitter, and YouTube have become valuable sources of data for understanding opinion dynamics due to their ability to track user behavior and infer opinions from content.

**Reddit.** Reddit[5] is a widely-used online platform where users can submit various types of content, such as text posts, links, images, and videos, and engage in discussions through comments. Reddit's attractiveness as a data source stems from several factors. Firstly, its popularity ensures a vast amount of data available for analysis. According to Similarweb Ltd. (September, 2023), Reddit ranks as the 18th most popular website globally. The platform is organized into "subreddits", which are topic-specific communities. Most subreddits are public, and even private ones can often be accessed upon request. Public subreddits can be viewed, commented on, and voted on by all registered users, and the data can be downloaded for free

[5]https://www.reddit.com/

by any registered user. With over 138,000 active subreddits, researchers can easily identify and study specific populations. Reddit is an anonymous environment that allows users to express their true beliefs, making it a valuable resource for studying sensitive topics. Researchers can access Reddit data through the site itself or its APIs. Reddit's official API is free and publicly available, offering a range of functions. Additionally, there are alternative ways to access Reddit data, such as Pushshift, a social media data collection, analysis, and archiving platform founded by Jason Baumgartner in 2015 [20]

**Twitter.** Twitter [6] is a microblogging platform where users can post short messages called *tweets* and interact with others through replies, retweets, and likes. Twitter provides a wealth of data that can offer insights into public opinion and behavioral responses in specific situations [39]. Researchers have been able – up until a few months ago [8] – to freely analyze tweet content, user interactions, and information dissemination to understand opinion dynamics on Twitter [2]. The platform's data availability and structure have contributed to its widespread success among researchers and third-party developers. Twitter data is freely available, public by default, primarily textual, and easily understandable. Twitter data can be categorized into two types: tweet-related information and user-specific information. Tweet-related information includes the textual content, time and location of production, and relational nature of the messages. User-specific information comprises the username, self-declared location, and lists of users followed and following the user. Despite its simplicity, this information can be combined to provide valuable insights into various aspects of Twitter usage, such as posting topics, strategies, and community formation and evolution.

Other platforms like Gab, Mastodon, Weibo, TikTok, and Instagram also offer unique opportunities for studying opinion dynamics and user behavior. Gab emphasizes free speech and user privacy; Mastodon is an open-source, decentralized social network; Weibo is a Chinese microblogging platform; TikTok is a short-form video-sharing platform popular among younger audiences; and Instagram is a photo and video-sharing platform that highlights the visual aspects of social interactions. Each platform provides distinct features, data accessibility, and research potential, making them valuable resources for understanding various aspects of online social interactions. By leveraging the wealth of data available on these platforms, researchers can gain valuable insights into the dynamics of opinions and behaviors in the digital age.

**Opinion dynamics models validated with real data.** Conclusions from different models appear to be realistic and seem to explain some real-world phenomena in a plausible way. However, there is no agreement on which models and characteristics better represent social interactions, and one of the reasons is that there is a scarcity of research in this direction. Outputs from discrete models have been compared to patterns seen in the data, such as voter model and election output data.

Recently, past voting records and the voter model were used to forecast election results in the US and UK [222]. In some cases, distributions seen in real data from surveys [147, 208, 134], social experiments [212], or social media [231] are used to tune parameters or modify agent-based models manually.

[6] Twitter has been recently rebranded to X [7]. For the sake of this thesis, we will continue to refer to it as Twitter to not create discontinuity in terminology throughout the work and with previous literature.

[8] https://www.engadget.com/twitter-shut-off-its-free-api-and-its-breaking-a-lot-of-apps-222011637.html

Agent-based models of opinion dynamics have the advantage of including causal mechanisms that make the models interpretable. However, they do not exploit the availability of data, and parameter calibration is a manual and difficult task. Some researchers have tried to tackle the opinion dynamics understanding problem in the last years using more empirical approaches. Some studies employ Bayesian learning techniques. Monti et al. in [163] proposed a learnable generalization of an opinion dynamics model [125] and tried to estimate the backfire effect and latitude of commitment of a political discussion on Reddit. This kind of approach maintains the causal interpretation possible while allowing for model selection and hypothesis testing on real data. In this study, the only observables considered are actions and interactions, while in [203], opinions are considered fully observable, and estimation of parameters through maximum-a-posteriori is used to find the most influential nodes. The approach is applied to Twitter and Reddit datasets. In [52], an ad-hoc model of opinion dynamics is developed, and then Bayesian inference is used to calibrate model parameters from real data. The model was further developed in [137], where each user is assigned a recurrent neural network to learn non-linearly from past timings and opinions.

# Part II

# Models for Biased Digital Environments

# Chapter 5

# Mass Media Impact on Opinion Evolution in Biased Digital Environments: a Bounded Confidence Model

Opinions and beliefs shape individual behavior, which drives human actions and a society's collective behavior, influencing politics, public health, and the environment. Changes in public opinion - even the formation of committed minorities - may profoundly affect decision-making and politics: a recent example is the temporary suspension of the Oxford-AstraZeneca vaccine during March 2021 [1], which has caused a slowdown in the vaccination strategy [22, 180], with possible direct consequences on public health. Therefore, to understand emergent collective behavior, it is desirable to understand better how these different factors interact to shape our opinions.

Why do we need models to understand opinion formation? We need them because the process of opinion formation, traditionally studied by social scientists or psychologists, is the result of the interaction of internal and external factors.

Social interactions [165] are the main ingredient driving the opinion evolution process. According to social influence theory [79], an interaction between social agents typically reduces the difference between their opinions or, at worst, leaves it unchanged. Besides social influence, opinion formation also depends on the information people collect from external sources (mainly in the form of mass media broadcasts), enhancing awareness of socio-political issues and events [95, 114]. For instance, traditional mass media have been argued to influence individual and public health [29, 225] on issues ranging from eating disorders [218], tobacco consumption [81], and vaccinations [200].

Besides information social agents can access, and how information is presented to them, a series of internal mechanisms play an important role in shaping opinions and beliefs. The way people process information is, in fact, far from being perfectly rational and is highly influenced by psychological factors and cognitive biases[2]. Psychological studies [126, 127] have observed that people, both online and offline, feel discomfort when encountering opinions and ideas that contradict their existing beliefs, i.e., experience cognitive dissonance [70]. Such cognitive biases have often been

[1] Reuters Institute (2021). *Suspension of AstraZeneca Shots Is 'Political Decision': Italy's Medicines Regulator Head.* https://bit.ly/3ki4Wk7

[2] Buster Benson. *Cognitive bias cheat sheet. An organized list of cognitive biases because thinking is hard. (2016)*https://betterhumans.pub/cognitive-bias-cheat-sheet-55a472476b18

studied through models of bounded confidence [54], i.e., the tendency to ignore beliefs that are too far from our current ones, or mimicking the backfire effects [175], i.e., the tendency to reject countering evidence and to strengthen the support to the current belief.

While such a dynamic has always existed, how people retrieve information has profoundly changed in the last twenty years. Television remains the most common media source among Europeans [3], but the use of the Internet and online social networks (OSNs) is steadily rising alongside the decline of the readership of newspapers.

However, OSNs are also environments where individuals express their opinions, discuss, and share content from other sources. These environments are ruled by algorithms that filter and personalize each user's experience accordingly to their and their friends' past behavior. This is intended to maximize users' engagement and enhance platform usage; however, it is theorized that filtering algorithms and recommender systems are likely to create an algorithmic bias [204]. By showing people only narratives aligned with their own existing beliefs, a positive feedback loop is obtained, reducing the amount of diversity in the user experience, contributing to the creation and maintenance of echo chambers [43] and filter bubbles [26, 181, 91]. Although personalization is essential in information-rich environments (to allow people to find what they are looking for and increase user engagement), there is great concern about the negative consequences of algorithmic filtering. Therefore, understanding how these different factors impact public opinion and how cognitive and algorithmic biases play a role in social influence mechanisms is essential to enrich our understanding of human behavior and also to define mitigation strategies to avoid unintended consequences.

In this Chapter, and in the following ones, we approach such a goal through the lens of opinion dynamics models [35, 205, 172, 171, 149, 59, 234, 183]. In particular we extend and study the *AB Model* [204] (which, in turn, extends the Deffuant-Weisbuch one [54]) to account for the role of external agents in a biased online environment (cf. Chapter 5), network effects (cf. Chapter 6) and adaptive network topology (cf. Chapter 7) with the possibility of higher-order interactions (cf. Chapter 7).

The content of this Part refers to 3 articles [177, 176, 179].

In the following, we extended [204], adding the possibility to specify a number of external mass media agents, defining the opinions they promote, and the frequency of agent-media interactions. We conducted numerical simulations to examine this extended model and analyzed the outcomes within the context of mean-field scenarios. In Sections 7.3 and 8.3, we will exploit the here-developed model and a methodology for the estimation of the confidence bound to perform a case study on a real-world network. We will see how this model – initialized with the real network structure and the initial opinion distribution inferred from the data – calibrated with a heterogeneous $\epsilon$ distribution estimated from real data effectively captures a behavior that the baseline model fails to capture.

The rest of this Chapter is organized as follows: in Section 5.1 we define the model and describe the performed simulations in greater detail, and we also describe the metrics used to analyze the results; in Section 5.2, we present and discuss the main results obtained from the simulations of the Algorithmic bias model with mass media; in Section 5.3 we sum up the work done, outline limitations and some future directions.

[3] Eurobarometer (2022). *Media & News Survey 2022.* https://europa.eu/eurobarometer/surveys/detail/2832

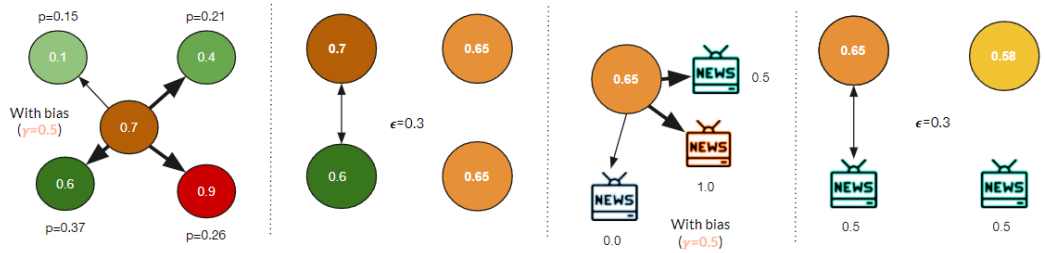



FIGURE 5.1: **Example of agent-to agent and agent-to-media interaction with $\gamma = 0.5$ and $\epsilon = 0.3$.** In the example, an agent with opinion 0.7 has a different probability of choosing one of the four neighbors, represented by the thickness of the arrows in the figure. After changing opinions due to the peer-to-peer interaction, the target agent chooses to interact with one of the three media, with a probability $p_m$. The choice of which media to interact with is determined according to $\gamma$, in the same way as in the social interaction: the higher the bias $\gamma$, the higher the probability of interacting with a media promoting a closer opinion to the current one of the agent. If the media falls within the agent's confidence bound $\epsilon$, the agent averages his opinion with the one of the media; otherwise, nothing happens. The media opinion, instead, remains unchanged.

## 5.1 Model and methods

To introduce in the study of opinion dynamics the idea of a recommender system generating an algorithmic bias, the classical *DW Model* [54] was previously extended in the Algorithmic Bias model (or *AB Model*, hereafter) [204]. Our work is an extension of the *AB Model* to include external information. The *DW Model* has been defined in Definition 11, while a definition of the *AB Model* can be found in Definition 18, both in Chapter 1.

To briefly recap, in the *DW Model*, we have a population of $N$ agents, where each agent $i$ has a continuous opinion $x_i \in [0, 1]$. At every discrete time step, a pair $(i, j)$ of agents is randomly selected, while in the *AB Model*, at every discrete time step, an agent $i$ is randomly picked from the population, while $j$ is chosen from $i$'s peers according to Definition 18:

$$p_i(j) = \frac{d_{ij}^{-\gamma}}{\sum_{k \neq i} d_{ik}^{-\gamma}}$$

where $\gamma$ indicates the bias strenght and $d$ the opinion distance between $i$ and $j$.

In both models, if the chosen agents' opinion distance is lower than a threshold $\epsilon$, $|x_i - x_j| \leq \epsilon$, then both of them change their opinion according to Equation (5.1):

$$\begin{aligned} x_i(t+1) &= x_i + \mu(x_j - x_i) \\ x_j(t+1) &= x_j + \mu(x_i - x_j). \end{aligned} \tag{5.1}$$

The *AB Model* introduces another parameter to model the algorithmic bias: $\gamma \geq 0$. This parameter represents the filtering power of a generic recommendation algorithm: if it is close to 0, the agent has the same probability of interacting with all its peers. As $\gamma$ grows, so does the probability of interacting with agents holding similar opinions while interacting with those who hold distant opinions decreases.

### 5.1.1 The Algorithmic Bias Model with Mass Media Agents

We now present our extension of the *AB Model*, tailored to analyze the effects of mass media propaganda. We chose to model mass media as stubborn agents connected to everyone in the population, i.e., agents whose opinions remain fixed during the dynamic process and can interact with the whole population. This choice simplifies real-world media outlets that may instead change the promoted point of view, being influenced by public opinion or politics. However, we assume that our analysis is temporally constrained and that such changes are unlikely. A completely mixed population model that every individual can use any media - offline and online - as an information source. The fact that individuals often have a limited set of sources among which they choose is due mainly to cognitive and technological biases, which we are trying to capture with this model. Finally, we allow an arbitrary number of media sources $M$ instantiated with custom opinion distribution $X_M$ to explore different scenarios in the present model. To regulate the interactions with media outlets, we added another parameter, namely $p_m \in [0,1]$, which indicates the probability that during each iteration of the model simulation - in addition to interacting with a peer - each agent interacts with a media $m \in M$.

**Definition 19 (Algorithmic Bias model with Mass Media *ABMM Model*)** *Let us assume a population of N agents, where each agent i has a continuous opinion $x_i \in [0,1]$ and a population of M mass media with fixed continuous opinions $x_{m_i} \in [0,1]$. At every discrete time step, an agent i is randomly picked from the population, while j is chosen from i's peers according to Definition 18. If their opinion distance is lower than a threshold $\epsilon$, $|x_i - x_j| \leq \epsilon$, then both of them change their opinion according to Equation* (5.1)*. With probability $p_m \in [0,1]$ agent i picks a mass media agent $m \in M$ according to Definition 18 (if $|M| > 1$). If $|x_i - x_{m_j}| \leq \epsilon$, agent i updates its opinion according to Equation* (5.1)*.*

Figure 5.1 illustrates an example of an interaction (both agent-to-agent and agent-to-media) and its effects on the node's opinion in the presented model.

To conduct our experiments, we implemented the *AB Model* with mass media within the NDlib [192] Python library. This library has many opinion dynamics, epidemic models, and a large user base. Adding our model to the library increases its availability to the scientific community.

### 5.1.2 Analyses and Measures

We simulate our model on a fully connected population of 100 agents, where the initial opinions are uniformly distributed, and we averaged the results over 100 runs. Like in [204], to avoid undefined operations in equation Definition 18, when $d_{ik} = 0$ we use a lower bound $d_\epsilon = 10^{-4}$. We imposed the simulations to stop when the population reaches an equilibrium, i.e., the cluster configuration will not change anymore, even if the agents keep exchanging opinions. We also set an overall maximum number of iterations at $10^6$ to account for situations where an equilibrium may never be reached. To better understand the differences in the final state, we studied the model for various combinations of the model parameters. We are interested in whether the different numbers and positioning of mass media and the growing interaction probability influence the final configuration, enhancing or reducing fragmentation and radicalizing individuals towards more extreme opinions, all other parameters being equal.

We replicated the work of [204] by setting a null probability to interact with the media to define a reliable baseline for comparison.

In the simulations, we evaluated the model on every combination of the parameters over the following values:

- $p_m$ takes values in $[0.0, 0.5]$, with steps of 0.1 - where for $p_m = 0$ the model becomes the *AB Model*.
- $\epsilon$ takes value in $[0.1, 0.5]$, with steps of 0.1.
- $\gamma$ takes value in $[0.5, 1.5]$, with steps of 0.25, and 0.0 - where for $\gamma = 0$ and $p_m = 0$ the model becomes the *DW Model*.
- $\mu = 0.5$, so whenever two agents interact, if their opinions are close enough, they update to the average opinion of the pair.

We analyzed different scenarios to understand the effects of (i) one media, either extreme with a fixed opinion of $x_{m1} = 0.0$ or moderate with an opinion of $x_{m1} = 0.5$, (ii) two extremist media with $x_{m1} = 0.05, x_{m2} = 0.95$ and (iii) two extremist media and a moderate one with opinions $x_{m1} = 0.05, x_{m2} = 0.5, x_{m3} = 0.95$.

**Measures**

We used different measures to interpret the results, each equally necessary to understand the final state of the population. The first and most intuitive measure to understand fragmentation is the number of clusters present on average at the end of the dynamic. We used a naive clustering technique to partition the final opinion distribution into clusters: we sorted the final opinions in each run and set a threshold. Starting from one extreme, the corresponding nodes belong to two clusters every time two consecutive opinions exceed the threshold. Optimal results were obtained using a threshold of 0.01. Once we divided the population into opinion clusters, we compute the cluster participation ratio, as in [204]:

$$C = \frac{(\sum_i c_i)^2}{\sum_i c_i^2} \tag{5.2}$$

where $c_i$ is the dimension of the $i$th cluster, i.e., the fraction of the population we can find in that cluster. In general, for $n$ clusters, the maximum value of the participation ratio is $n$ and is achieved when all clusters have the same size. At the same time, the minimum can be close to one if one cluster includes most of the population and a tiny fraction is distributed among the other $n$ min 1.

To grasp the attractive power of the media in each setting, we also computed the number of nodes present in the clusters centered on the media opinion. Specifically, we consider the percentage of agents that hold opinions in the range $[x_m - \lambda, x_m + \lambda]$ with $x_m$ being the media opinion and $\lambda = 0.01$.

## 5.2 Results

The present work aims to extend the Algorithmic Bias model [204] to understand how interacting with mass media in a biased environment (i.e., ruled by recommender systems and filtering algorithms) influences the outcome of the opinion evolution. In our simulations, we consider 100 agents with continuous opinions in the interval $[0, 1]$, which can model opinions on any issue, with values 0 and 1 representing the most extreme opinions. The agents are allowed to interact with each other at discrete time intervals and with a fixed number of $M$ stubborn agents,

representing traditional media outlets that promote a fixed opinion over the whole time period. To represent this environment realistically, interactions (agent-to-agent and media-to-agent) are subject to cognitive and algorithmic biases. The stronger the algorithmic bias, $\gamma$, the higher the probability of interacting with similar agents and the lower the probability of interacting with different ones. Cognitive bias - specifically bounded confidence - limits interaction to an agent's opinion neighborhood: two agents influence each other (according to social influence theory, adopting their mean opinion) if and only if their initial opinion distance is below a certain threshold $\epsilon$. This parameter is constant across the whole population and over time. In the remainder of the present work, we often refer to it as the level of "open-mindedness" of the population because bounded confidence and open-mindedness both involve a willingness to consider different perspectives within certain limits. On the other hand, influenciability refers to being easily swayed by others, regardless of the strength of their arguments. Thus, we felt that open-mindedness was a more appropriate term for describing the bounded confidence threshold in our work (for example, as in [196]). However, it's important to note that in opinion dynamics models, behavioral and psychological factors are often simplified and represented by model parameters. As a result, nuances can be lost, and the bounded confidence threshold could also be interpreted as influenciability. To control the frequency of interactions with the media, we set a fixed probability $p_m$ - constant over time and across the whole population - which defines how likely it is to interact with a news piece (stubborn agent) after a user-to-user interaction. In our experiments, we assumed a mean-field context (e.g., all individuals can interact with all other agents without any social restrictions), which is a good starting point for analyzing the behavior of an opinion dynamics model. The model is detailed in Section 5.1.

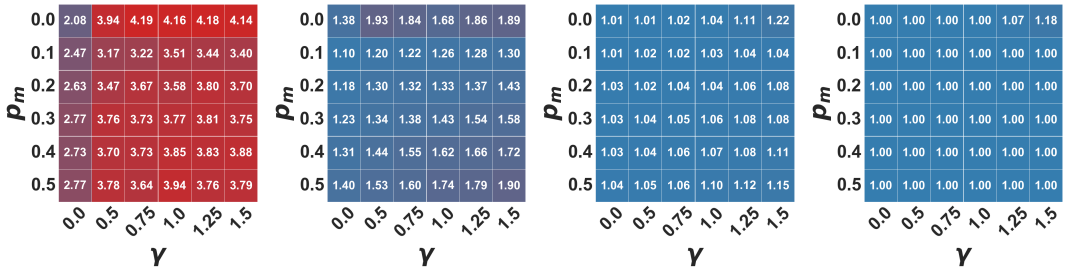
The scenarios we analyzed in the present work are (i) a single moderate media ($x_m = 0.5$), to discover whether a "moderate message " would prevent the population from polarising in cases where it would happen without propaganda; (ii) extremist propaganda, where there is only one news source promoting a fixed extreme opinion (in this case, it was set to $x_m = 0.0$, but the same conclusions hold for 1.0); (iii) two polarised media sources, promoting two opinions at the opposite sides of the opinion spectrum ($x_{m1} = 0.05$ and $x_{m2} = 0.95$); (iv) finally, we also investigated a more balanced scenario where there are two polarised media sources (same as above) and a moderate one (promoting the central opinion of the spectrum, i.e., $x_{m3} = 0.5$).

Without external effects, the population tends to: (i) polarise around moderately extreme positions (i.e., 0.2 and 0.8) when agents are "close-minded" ($\epsilon \leq 0.32$); (ii) reach consensus around the mean opinion (i.e., 0.5) when agents are "open-minded" ($\epsilon > 0.32$), while the recommender system increases polarization/fragmentation, as shown in [204].

In the remainder of this section, we analyzed these four different media landscapes and their effects on the opinion dynamics compared to the baseline model [204].

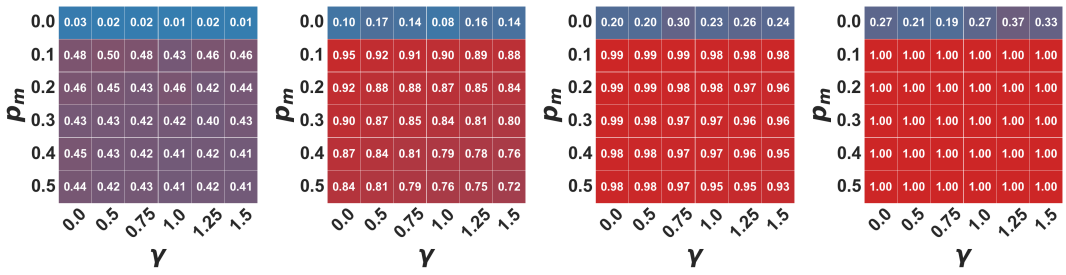
### 5.2.1 A moderate media in a biased environment favors the emergence of extremist minorities

In the first setting, we analyzed the effects of a "moderate message" on the opinion formation process, i.e., a single mass media promoting a central opinion ($x_m = 0.5$). We start from the hypothesis that such a media landscape may counteract the polarizing effects of a low bounded confidence $\epsilon$ or the fragmenting effects of a high




(A) $\epsilon = 0.2$ (B) $\epsilon = 0.3$ (C) $\epsilon = 0.4$ (D) $\epsilon = 0.5$

FIGURE 5.2: **Average number of clusters in the moderate setting.** In the figure, the average number of clusters of the final opinion distribution is represented as a function of the algorithmic bias $\gamma$ and the probability of user-media interaction $p_m$ for different bounded confidence values $\epsilon$. Values are averaged on 100 independent runs of each setting.




(A) $\epsilon = 0.2$ (B) $\epsilon = 0.3$ (C) $\epsilon = 0.4$ (D) $\epsilon = 0.5$

FIGURE 5.3: **Average percentage of agents in the media cluster (0.5) in the moderate setting.** In the figure, the average percentage of agents in the moderate cluster (0.5 +- 0.01) of the final opinion distribution is represented as a function of the algorithmic bias $\gamma$ and the probability of user-media interaction $p_m$ for different bounded confidence values $\epsilon$. Values are averaged on 100 independent runs of each setting.

algorithmic bias $\gamma$. Bounded confidence, as in the baseline model, can be so high that all agents are eventually drawn towards the same opinion (regardless of the strength of algorithmic bias), as in the case of $\epsilon = 0.5$ (Figure 5.2(D)). In general, in this setting, both cognitive and algorithmic biases maintain the effects they have in the baseline model: a higher confidence bound is more likely to push the population towards consensus, while a higher algorithmic bias increases the level of fragmentation in the final opinion distribution.

What emerged from our simulations is that, when interactions are not mediated by the recommender system ($\gamma = 0$), **fragmentation** increases with the frequency of agent-to-media interactions: in fact, the average number of opinion clusters at equilibrium (see Figure 5.2) increases with ($p_m$). Such tendency is due to the fact that, by increasing $p_m$, the portion of the population that initially has the media within their confidence bound moves towards such opinion faster than in the baseline model, thus reducing the probability of attracting agents at a distance greater than $\epsilon$ from the media that, in turn, will eventually stabilize around more extreme positions. When the social dynamic is, instead, mediated by a filtering algorithm, biasing the choice of the interacting partner towards like-minded individuals, the level of opinion fragmentation in the population is initially lower (for small $p_m$) with respect to the baseline model ($p_m = 0.0$), but - likewise - it grows as agent-to-media interactions

become more frequent. These results disprove our initial hypothesis that a "moderate" propaganda may straightforwardly counter polarization/fragmentation. Instead, promoting a single "moderate" opinion may not push the population to conform towards the desired point of view. Fragmentation is reduced only when the frequency of interaction with media is low. Otherwise, it also becomes a fragmenting factor.

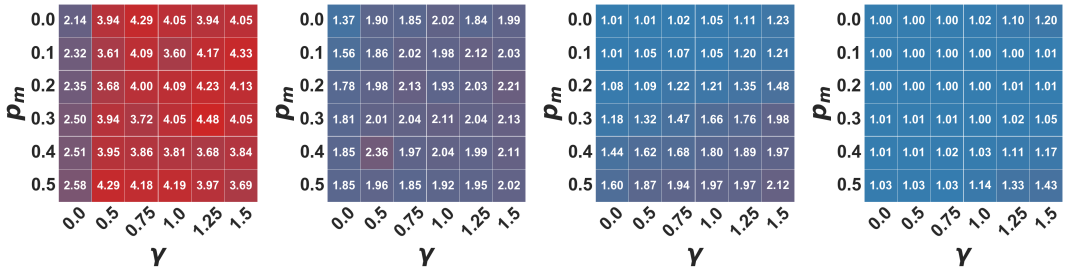
Besides the number of clusters that coexist in the stable state, if we look at the whole opinion evolution process, we can see that there is always a portion of the population clustering around the media opinion (i.e., with opinion $x_i \in [0.5 + / - 0.01]$, while a small fraction assumes extremist positions. Figure 5.3 shows this cluster's population percentage. The more open-minded the population and the higher the frequency of agent-to-media interactions, the larger the portion of agents that the media can rapidly attract towards the average opinion: thus, pushing the population towards consensus and countering the slowing down effect created by the algorithmic bias. Moreover, as we can see from Figure 5.3, while in the baseline model, only a narrow portion of the population assumes the mean opinion when a moderate media is promoting that opinion, we can see that the portion of the population ending in the moderate cluster in the steady state grows even with just a low probability to interact with the media and narrow open-mindedness threshold. Therefore, while consensus is not fully reached, a major cluster around the media is observed. Conversely, in the case of media absence ($p_m = 0.0$), there is a higher variability in the final size of the moderate cluster. Even when a consensus forms, it is not necessarily around the mean opinion. Otherwise, the population polarizes around mildly extreme ones (around 0.2 and 0.8), avoiding the creation and maintenance of strongly extremist minorities, as it happens in the present model.

However, when interactions are mediated by a filtering algorithm - $\gamma > 0$, the media can attract a smaller fraction of the population since agents holding more extreme opinions are much less likely to interact with those in the sphere of influence of the moderate media. Overall, our experiments showed that the algorithmic bias maintains its fragmenting power: specifically, as the bias grows, the extremist clusters that coexist with the moderate one increase in size but also in dispersion, eventually splitting into multiple smaller clusters. At the same time, the fragmenting effect of the recommender system decreases the size of the moderates/neutrals cluster, especially in the case of moderately close-minded populations (Figure 5.3), but not in a significant way (at least with the population size considered in the present work).

### 5.2.2 Extremist media shifts consensus in open-minded populations

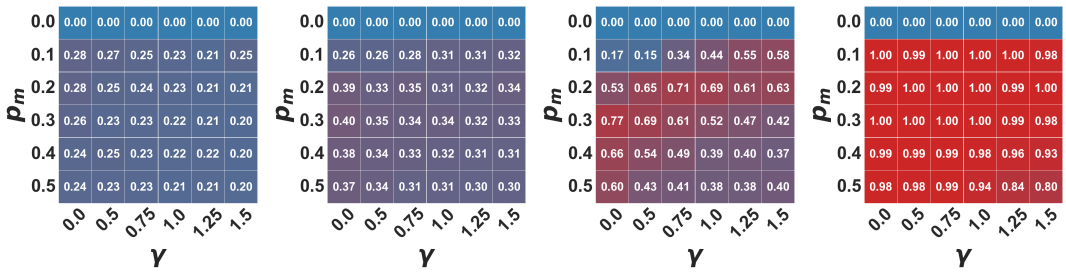
To investigate the effects of extremist propaganda and its effectiveness in shifting the consensus towards the desired opinion, we set the number of mass media outlets to $M = 1$ and the promoted opinion to $x_m = 0.0$.

Like in the moderate setting, the baseline model's cognitive and algorithmic biases effects also remain in this setting. In the same way, an increase in the frequency of interaction with extremist propaganda (when $\gamma = 0$) translates into an increase in the fragmentation of the final population. The number of clusters of the final opinion distributions, in fact, grows with $p_m$ (Figure 5.4). For example, when the population is close-minded ($\epsilon = 0.2$), in the absence of propaganda ($p_m = 0$), in the final state, there are two main clusters (on average), while as $p_m$ increases, the number of clusters approaches 3. In the same way, as the population is more "open-minded" - so




(A) $\epsilon = 0.2$ (B) $\epsilon = 0.3$ (C) $\epsilon = 0.4$ (D) $\epsilon = 0.5$

FIGURE 5.4: **Average number of clusters in the extremist setting.** In the figure, the average number of clusters of the final opinion distribution is represented as a function of the algorithmic bias $\gamma$ and the probability of user-media interaction $p_m$ for different values of $\epsilon$. Values are averaged on 100 independent runs of each setting.




(A) $\epsilon = 0.2$ (B) $\epsilon = 0.3$ (C) $\epsilon = 0.4$ (D) $\epsilon = 0.5$

FIGURE 5.5: **Average percentage of agents in the media cluster (0.0) in the extremist setting.** In the figure, the average percentage of users in the extremist cluster ([0.0, 0.01]) is represented as a function of the algorithmic bias $\gamma$ and the probability of user-media interaction $p_m$ for different values of $\epsilon$. Values are averaged on 100 independent runs of each setting.

the number of clusters in the baseline model is lower - interacting with the propaganda still generates an increase in the number of clusters (moving the population from consensus around one opinion to clustering around two opinion values for $\epsilon = 0.3$ and also $\epsilon = 0.4$, even if in this case on average there is a consistent majority cluster). Despite the fact that an extreme opinion is promoted (while, without external effects, agents tend to conform to moderate positions), in this case, bounded confidence or, in other words, the level of "open-mindedness" of the population, can be so high that all agents are eventually drawn towards the same opinion, as in the case of $\epsilon = 0.5$ (Figure 5.4(D)). This fact still holds when the interactions are mediated by a recommender system ($\gamma > 0$), biasing the choice of the interacting partner towards like-minded individuals, but it is less evident due to the fragmenting power of the algorithmic bias. For example, when the population is close-minded, we tend to have an average of three or four clusters in a biased environment.

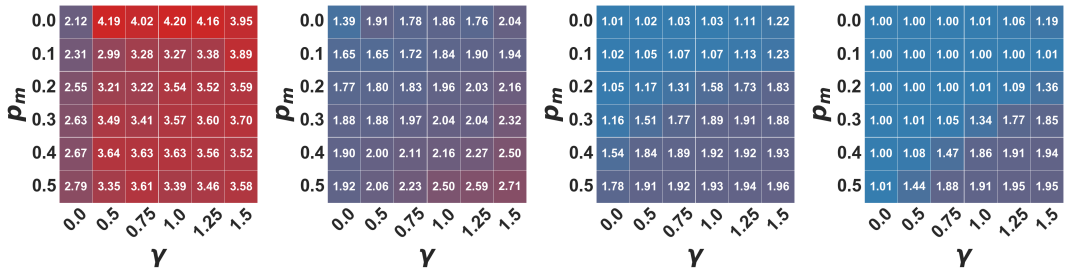
It is important to note that, compared to the moderate situation, the fragmenting effect of the external media is stronger for an extremist message. The number of clusters reported in Figure 5.2 is generally smaller than that reported in Figure 5.4.

In the present model, differently from the baseline [204], i.e., $p_m = 0.0$, the population splits into more than one cluster when $\gamma > 0$ and $\epsilon$ is sufficiently low. One of these clusters always forms around the extreme media opinion ($x_m = 0.0$) while - as the bias grows - the rest of the population either clusters around a single value on the

opposite side of the opinion spectrum or fragments into multiple small clusters (and their distance from the extremist propaganda increases with the open-mindedness of the population). This effect is stronger as the algorithmic bias increases and as the frequency of interaction with the media grows. In the case of extremist propaganda, as we can expect, a higher portion of the population in the stable state is an extremist, holding the same opinion promoted by the media (see Figure 5.5). Additionally, the higher the open-mindedness of the population, i.e., the higher the confidence bound $\epsilon$, the higher the dimension of the extremist cluster - until ($\epsilon \geq 0.5$) the population is entirely attracted towards this extreme position (Figure 5.5(D)). However, as the bias increases, the final number of opinion clusters increases, and the average number of agents in the extremist cluster decreases: the fact that algorithmic bias increases fragmentation in the population causes - in this case - the formation and maintenance of an "opposition" cluster, countering the process of complete radicalization of the population. As the bias increases, of course, this cluster becomes more dispersed with respect to its average opinion, and for extreme biases, it fragments into a series of small opinion clusters. Therefore we can conclude that algorithmic bias acts as a partial protector against the message of one extremist media.

It is also worth noticing that, with $p_m > 0$, all other parameters being equal, the size of the extremist cluster does not increase with the probability of interaction with the media; on the contrary, the maximum size is reached for low or intermediate values of $p_m$ (see Figure 5.5). Also, in this case, such behavior is tied to the fact that even if the frequency of interaction with the media increases, those agents that initially are within the sphere of influence of the media will converge towards an extremist position more rapidly, thus losing the ability to attract those who are outside of it. When dealing with close-minded agents, less frequent propaganda can attract a higher fraction of the population with respect to more intense propaganda. If the population is open-minded, the frequency of interactions with the media loses most of its discriminant power: if at least half of the agents are already initially influenceable by the media, the whole population will converge toward the media opinion.

### 5.2.3 Polarised media increase the divide




(A) $\epsilon = 0.2$ (B) $\epsilon = 0.3$ (C) $\epsilon = 0.4$ (D) $\epsilon = 0.5$

FIGURE 5.6: **Average number of clusters in the polarised setting.** In the figure, the average number of clusters of the final opinion distribution is represented as a function of the algorithmic bias $\gamma$ and the probability of user-media interaction $p_m$ for different values of the cognitive bias $\epsilon$. Values are averaged on 100 independent runs of each setting.

Public debates are often characterized by bi-polarity, a situation where two opposing views are proposed and debated. For example, media polarization in the U.S. has increased in the past half-decade, and both liberal and conservative partisan media are likely contributing to polarization in the Cable news networks [153]. While

acknowledging that our synthetic setting represents a simplification of the complex dynamics at play, it nevertheless presents a scenario that merits further investigation.

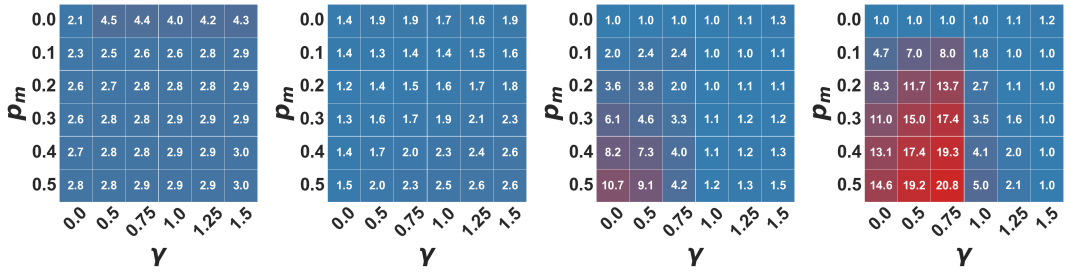
To recreate such a scenario - even if simplistically -, we simulated the presence of two extremist media outlets in the population, promoting opinions at the opposite sides of the opinion spectrum, - i.e., we set $x_{m1} = 0.05$ and $x_{m2} = 0.95$. As expected, the presence of two polarised media increases the system's polarization, which would already naturally arise due to the effects of the cognitive and algorithmic biases ($\epsilon \leq 0.3$), but the presence of the media pushes the population towards the media opinions - which are more extreme than the ones that naturally form in the baseline model (see Figure 5.6(A)-(B)). The presence of these two media, moreover, can bring the population towards polarization/fragmentation even in cases where the baseline model would predict full consensus ($\epsilon = 0.4$), a fragmentation exacerbated by the recommender system effects (see Figure 5.6(C)-(D)). On the other hand, in "close-minded" populations, the byproduct of agent-to-media interactions increasing the number of opinion clusters is that the rapid polarization of the extremes of the population results in the formation of a cluster of "moderate" agents, coexisting with polarized groups. On the one hand, this reduces the level of polarization in the population with respect to the baseline model. On the other hand, the polarized subpopulations are more extremist than in the baseline. As the filtering power of the recommender system increases, such a moderate cluster splits into multiple small ones, still concentrated around the center of the opinion spectrum. Moreover, as the algorithmic bias grows, the two extremist clusters reduce their sizes, and more agents become neutral, even if they hold a wider range of opinions. This is because a reduced fraction of agents interacts with extremist media and/or peers that end up in the extremist cluster early in the process. Therefore, they cannot attract a more significant portion of the population with respect to the case where the filtering power of the recommender system is more robust. As the open-mindedness of the population grows, an increasingly stronger algorithmic bias is needed to maintain the moderate cluster, and, in most cases, the population tends to polarise, with the two sub-populations approaching the media opinions. The population is, in this scenario, ultimately radicalized around very extreme positions (0.05 or 0.95), similar to the case of a single extreme media. Finally, the recommender system makes the polarization process faster than what was observed in the baseline model, allowing fewer opinion clusters to coexist during the opinion dynamics.

### 5.2.4 Open-minded populations are unstable in a balanced media landscape

In the last setting, we considered a more balanced information environment, with the presence of two extremist media in the population, promoting opinions at the opposite sides of the opinion spectrum, - i.e., we set $x_{m1} = 0.05$ and $x_{m2} = 0.95$, alongside with a moderate media, with $x_{m3} = 0.5$.

In this setting, agents can retrieve from mass media both moderate and extremist points of view.

This more balanced news diet appears to foster fragmentation still. In fact, the higher the frequency of agent-to-media interactions, the more fragmented is the final population, as we can see from the average number of opinion clusters in the final population, which grows with $p_m$ (Figure 5.7) and from the average pairwise distance, indicating how far are the peaks in the final opinion distribution.




(A) $\epsilon = 0.2$ (B) $\epsilon = 0.3$ (C) $\epsilon = 0.4$ (D) $\epsilon = 0.5$

FIGURE 5.7: **Average number of clusters in the balanced setting.** In the figure, the average number of clusters of the final opinion distribution is represented as a function of the algorithmic bias $\gamma$ and the probability of user-media interaction $p_m$ for different $\epsilon$ values. Values are averaged on 100 independent runs of each setting.

In this case, the algorithmic bias maintains its fragmenting power for a close-minded population (i.e., $\epsilon \leq 0.3$). As the bias grows, the number of clusters increases, but it never exceeds three (Figure 5.7(A)-(B)) since the population tends to converge towards the media opinions rapidly. The combination of a higher frequency of agent-to-media interactions, and the fact that interactions are biased towards similar opinions, allows each media to rapidly attract a portion of the population towards the promoted opinion.

On the other hand, in open-minded populations, $\epsilon \geq 0.4$, the relationship with the bias changes: from our experiments, it emerged that fragmentation is higher for low (Figure 5.7(C)) or intermediate (Figure 5.7(D)) values of the algorithmic bias $\gamma$, as the number of clusters in the final opinion distribution shows.

However, due to a stronger bias, the fragmentation that arises in the final state is not like the one reached in [204]. In that case, it was a stable state. In this case, the dynamic never reaches equilibrium, and agents keep changing their opinions influenced by the fixed opinions of the media.

Nevertheless, in the cases where consensus can be reached, if open-mindedness is high, the dynamic is still unstable, and it takes a long time for the population to reach a consensus. Let us recall that the distance between two adjacent media is 0.45, so when $\epsilon = 0.4$ agents holding an opinion between 0.10 and 0.45 or between 0.55 and 0.9 can be attracted by the moderate media and one extremist media that falls within their confidence bound, and this generates an unstable stationary state preventing the system from reaching equilibrium. Obviously, the higher the open-mindedness, the higher the number of clusters (and the average entropy of the final distribution) since agents are distributed on a wider opinion spectrum, and real clusters do not form. This effect is counteracted by a high algorithmic bias, which practically impedes the interaction with the furthest media, even if in the range of the confidence bound.

## 5.3 Discussion and Conclusions

A bounded confidence model of opinion dynamics with algorithmic bias and mass media agents was presented and studied in a mean-field setting. The model is an extension of the Algorithmic Bias model [204] – defined in Definition 18 in the present thesis – to include one or more mass media outlets. In the present work, media are

modeled as stubborn agents, each promoting a fixed opinion and connected to every agent of the population. We analyzed four different settings, each representing a specific media landscape: in the first, a single moderate media is present; in the second, the single media supports extremist propaganda; in the third, two polarised media promote extreme and opposite opinions; and in the latter, a third media, promoting a moderate opinion, is added to the polarised setting.

Our experiments reveal that mass media have an essential role in pushing people towards conformity and promoting the desired point(s) of view, but not in a straightforward manner, as adherence to the media message depends highly on cognitive and algorithmic bias and on the strength of the media itself. As we saw in the "moderate setting" (Section 5.2), an open-minded population tends to conform to moderate opinions, and only a few individuals will not. The main result of the "moderate message" is concentrating the central consensus cluster around the desired value. As expected, the size of the non-conforming clusters increases with algorithmic bias and decreases with open-mindedness. However, the size of the extremist nonconforming clusters also appears to increase in the strength of the moderate message. This is counterintuitive and indicates that, in general, not only the message has to be moderate, but also the frequency with which the message is presented has to be reduced. Moderation is necessary for all aspects to maximize adherence to the message.

Analyzing the results of the "extremist propaganda", we saw that the power to push individuals towards the media opinion is not dependent on such opinion. In this case, the open-minded population tends to become extremist because agents are pushed toward the media opinion and conform to that value. Again, we observe that the maximum adherence to the media message is always obtained for moderate frequencies of interaction with media.

In a polarised media landscape, with two poles promoting extreme and opposite opinions, the more "open-minded" is the population - or, in other words, the easier it is to change peoples' minds - the more likely the population will end up in one or two (oppositely) polarised extremist clusters. Also, in such a scenario, even when there would be a consensus around a moderate opinion, a higher frequency of interaction with the two extremist media is enough to push the population towards polarised stances, with two clusters forming around the media opinions.

In a balanced media landscape, when populations are close-minded, the more agents interact with mass media, the more they attract a portion of the population towards the promoted opinion. The effects of cognitive biases, i.e., bounded confidence, generally maintain the same role they have in the baseline model: the more "open-minded" is the population, the easier agents conform around the promoted opinion(s). However, when agents have access to multiple information sources (besides their peers' opinions), "open-mindedness" leads to a population of indecisive individuals and unstable dynamics that prevent the system from reaching equilibrium.

We typically give a positive value to a highly open-minded population, i.e., a population where agents have a high confidence bound. However, a higher open-mindedness in the presence of mass media may mean that the whole population is attracted to an extremist position, as we saw in the case of extremist propaganda or two polarised media. Even if the media is not extremist - it still means that the population conforms towards a single point of view, converging faster and perfectly towards a single opinion value, making agents subject to external control by those who can manipulate the information delivered by the media. Similarly, we usually give a positive value to the final consensus setting. However, as we already said,

consensus also means conformity, homologation to a standard, which may be imposed from the outside and manipulated through media control to achieve the goals of those in power and hardly the optimal situation for our societies and democracies.

The large amount of research that has focused on detecting the strength and the effects of recommender systems and algorithmic biases moves from the idea that the presence of such biases traps users into echo chambers and/or filter bubbles, preventing them from getting confronted with a balanced information diet and thus polarising/fragmenting the population into a series of opinion clusters that do not communicate. Even though this is still far from being proven, even if we assume that this effect is true, it is worth asking ourselves whether this always has a negative effect. For example, from our work, it emerged that the presence of a recommender system alongside a moderate message facilitates the emergence and maintenance of extremist minorities, which coexist with a group of moderates. However, both a lower confidence bound, $\epsilon$, and a higher algorithmic bias, $\gamma$, when acting in a context where there is extremist propaganda or two polarised extremist media, avoid the complete radicalization/extremization of the whole population and counter the complete polarization by favoring the presence of a moderate cluster in both cases. We also observed that the recommender system facilitates convergence in a balanced setting where the population is open-minded. Indeed, it prevents the dynamic from being completely unstable - i.e., avoiding agents continuously changing their opinion and never reaching a stable state due to the presence of conflicting sources.

The present work is a preliminary step toward analyzing the interplay of social and media influence in digital environments and presents several limitations. We focused on mean-field scenarios, which prevents us from considering possible network effects on the results of the opinion evolution process. While this is a sound starting point, the obtained insights must be tested against different network structures or real networks to employ the proposed model to analyze and understand reality fruitfully. Moreover, social connections change in real settings, influencing subsequent interactions and opinion exchanges. As we will see in Chapters 6 and 7 for the Algorithmic Bias model without mass media, network effects should be taken into account: greater sparsity in the underlying network structures appears to promote polarization and fragmentation in the Algorithmic Bias model, and it is possible that a similar effect may be observed in the model presented in this study.

We will also see in Chapter 6 that mesoscale structure may promote different outcomes on the dynamic based on the different initial conditions. Here, we studied this model on a real network that exhibits two polarized communities. Experiments suggest that this may favor consensus even for lower confidence threshold levels. In order to verify this hypothesis, more convergence analysis needs to be performed on different modular networks and with different initial conditions. The present model could then be studied on adaptive network topologies to understand the interplay of the dynamics on/of the network. Moreover, in our work, bias has a role in the choice of the media only when in the presence of two or more sources. Even in the presence of a single externally promoted opinion, some agents who are too far away from that position may still have a small probability of interacting with it. To account for such a pattern, the probability of interacting with the media - which is now homogeneous across the whole population - could be made heterogeneous and dependent on the distance between the agent's opinion and the promoted opinion and heterogeneous levels of agent engagement with mass media can be integrated within the model.

Although all the different models demonstrate that an open-minded population can reach a consensus on all issues, it is an unrealistic assumption. Regardless of how open-minded they may be, each user will still have an inherent preference towards

one side of the opinion spectrum. To address this, we propose extending the current model to incorporate a baseline opinion that consistently influences the user in that direction. Finally, as we will see in Chapter 8, real populations may have heterogeneous (opinion-dependent) levels of "open-mindedness", which could be taken into account to specify agents' peculiarities better, as well as heterogeneous activity levels as in [146]. Similarly to "open-mindedness" and activity levels, we plan to augment the current model with data-driven insights on media bias and user interactions with mass media and authoritative voices via online social networks. This will enable us to understand better the long-term impact of such interactions and how their influence differs from that of peers. One missing aspect in this context is undoubtedly a "dynamic" behavior from users, including the creation/destruction of links and the evolution of $\epsilon$ and $p_m$ with increasing/decreasing polarization. Additionally, there needs to be more evolution in the media's behavior or a more realistic user-media relationship. The media should be aware of the cognitive biases of their users, and not all media outlets have the entire population as their audience. The more polarized the media are, the more likely they are followed by only a portion of the already aligned population, thereby promoting ideas aligned with that population segment. Another aspect not considered is that in a real setting, the "media" or stubborn agents may not be mainstream media with which everyone can interact but specific influential users within the network. This model would need to be adapted in such a scenario, considering that these stubborn agents are no longer connected to the entire population but only to certain nodes. Furthermore, the nodes they are connected to might depend on the opinions of those nodes and the opinions they promote. While our model has some drawbacks, as discussed above, it also has some advantages: it is simple, it can be tested on various topologies, it considers psychological, technological, and external factors, and it allows for flexibility in the number and opinions of the media.

# Chapter 6

# The role of different network structures

In network science, connectivity is established when there is a path connecting any two nodes in a network, which is crucial for the proper functioning of network-based services, including communication networks. This particular property is also observed in online social networks despite their large size and sparseness, demonstrating the emergence of a massive connected component. According to the Erdős–Rényi random network model [64], a giant connected component emerges when the average network degree is 1. Real networks also display the small-world phenomenon [226], leading to a short average path length between two randomly chosen nodes that is proportional to the logarithm of the system size. Consequently, in a city with roughly 100,000 inhabitants, any two individuals are connected in just 3 or 4 steps. Real networks exhibit hubs, which are nodes significantly more connected than others. The Barabási–Albert model [15] explains this scale-free degree distribution by attributing the presence of hubs to the network's growth and preferential attachment. This phenomenon can lead to unequal resource distribution, as seen in social media, market success, and protein-to-protein interaction networks. In socio-economic scenarios, this could cause unsustainable levels of inequality, hindering the attainment of desirable social outcomes for society. Additionally, network clustering, as measured by the local clustering coefficient, is a common property of real networks, which expect higher clustering than the Erdős–Rényi model (for a more detailed explanation on networks and their properties, see Chapter 2 and Section 2.1).

Real networks have therefore an innate tendency to polarize and segregate, as demonstrated by their intrinsic characteristics. In digital environments, such a tendency is further exacerbated by AI-powered tools that, using big data as fuel, make personalized suggestions to every user to make them feel comfortable and, in the end, maximize their engagement [181]. Even if this kind of suggestion can be beneficial for a user at the individual level, from a societal point of view, it can lead to alarming phenomena in a wide range of domains.

In Chapter 5, despite accounting for an additional element of realism, i.e., the presence of mass media influencing the agents, we still considered a fully connected population, disregarding possible network effects on the dynamics.

In this Chapter and the following, we investigate the expected effects of the interplay between AI-powered tools (leading to algorithmic biases) and the emergent properties of underlying structures. In particular, moving from the results discussed in [204] where a mean-field context is assumed (e.g., all individuals can interact among them without any social restrictions), in the present, we aim to study the effect that different network topologies have on opinion formation and evolution when in the presence of a filtering algorithm.

Our goal is to verify if/how networks' structure exacerbates the polarization and fragmentation generated by the cognitive and algorithmic bias's presence. We want to verify if moving from a complete network with $L_{max}$ links to a network with $L << L_{max}$ links and a predetermined topology influences the final simulation state, making it harder for the population to reach a consensus or ultimately preventing it.

To such extent, and to allow results reproducibility, we focus our analysis on well-known network models, namely Erdős–Rényi [64] (to capture sparsity and the small-world phenomenon), Barabási–Albert [15] (for the role of hubs and scale-free degree distribution) and Lancichinetti–Fortunato–Radicchi benchmark [140] graphs (henceforth referred to as LFR graphs). Our research aims to finally investigate whether, in realistic environments, opinions remain trapped inside communities or not and which are the effects of different topologies on the steady state of the modeled dynamic process, e.g., whether they facilitate/ counteract polarization/ fragmentation or promote consensus. For a thorough description of these generative network models characteristics, please refer to Chapter 2 and Section 2.1.1.

Adopting such controlled environments, used to simulate the social structure among a population of interacting individuals, we analyze the behaviors of the Algorithmic Bias model [204] and discuss the role of graph properties on the observed simulation results.

## 6.1 Experimental analysis and results

In all scenarios, we set the number of nodes $N = 250$. For the ER network, we fix the $p$ parameter (probability to form a link) to 0.1 (thus imposing a *supercritical* regime, as expected from a real-world network); we obtain a random network composed of a single giant component with an average degree of 24.94. In the BA network, we set the $k$ parameter (number of edges to attach from a new node to existing nodes) to 5, thus creating a network with an average degree equal to 9.8.

We generated nine different networks using the LFR benchmark ($N = 250$). The parameters used for its construction have been set as follows:

- power-law exponent for the degree distribution, $\gamma = 3$;
- power-law exponent for the community size distribution, $\beta = 1.5$;
- fraction of intra-community edges incident to each node, $\mu_{LFR} \in \{0.1, 0.5, 0.9\}$;
- average degree of nodes, $< k >$ = 10;
- minimum community size $min_s = 50$, thus losing the power-law community size distribution and generating 4 communities of similar sizes in the end.

The parameter $\mu_{LFR}$ controls the number of edges between communities, thus reflecting the network's amount of noise. Therefore, the network with $\mu_{LFR} = 0.1$ has better-defined communities than the one generated with $\mu_{LFR} = 0.9$.

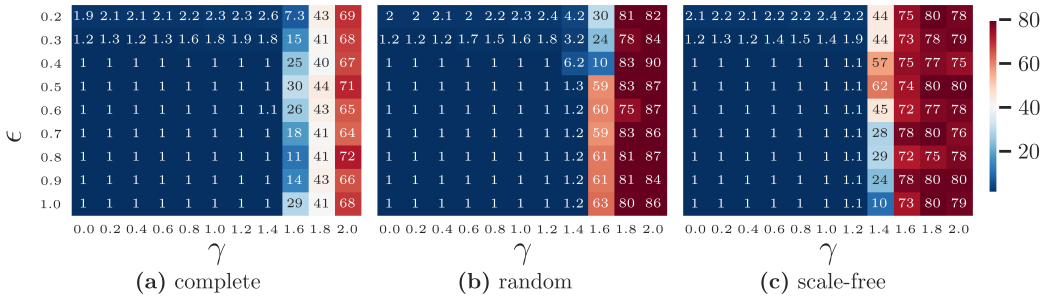
Like in [204], to avoid undefined operations in equation Definition 18, when $d_{ik} = 0$ we use a lower bound $d_\epsilon = 10^{-4}$. The simulations are designed to stop when the population reaches an equilibrium, i.e., the cluster configuration will not change anymore, even if the agents keep exchanging opinions. We also set an overall maximum number of iterations at $10^5$. To account for the model's stochastic nature, we compute the average results over 10 independent executions for each configuration, where the initial opinion distribution is always drawn from a random uniform

probability distribution in [0,1]. To better understand the differences in the final state concerning the different topologies considered, we study the model on all networks for different combinations of the parameters. We are interested in whether, parameters being equal, the different topology influences the final cluster configuration enhancing polarization and fragmentation, but also the dynamics of the process, by slowing down the convergence or reducing the density of the final opinion clusters. In the simulations, we tested the model on every possible combination of the parameters over the following values:

- $\epsilon$ takes a value from 0.2 to 1.0 with the step of 0.1.
- $\gamma$ takes value from 0 to 2.0 with the step of 0.2; for $\gamma = 0$ the model becomes the *DW Model*.
- $\mu = 0.5$, so whenever two agents interact, they update to the pair's average opinion if their opinions are close enough.

For the simulations of the *AB Model* on the LFR benchmark networks, instead, we tested the model over the following values:

- $\epsilon \in \{0.2, 0.3\}$. We impose this choice because in the mean-field, for these values, the number of clusters grows with increasing gamma, and we obtain a situation of polarization and fragmentation:
- $\gamma \in \{0.0, 0.5, 1.0, 1.5, 2.0\}$;
- $\mu = 0.5$. With this value, whenever two agents interact, if their opinions are close enough, they update to the pair's average opinion.




(a) complete (b) random (c) scale-free

FIGURE 6.1: **Average number of clusters across topologies.** The figure displays the average number of clusters as a function of $\epsilon$ and $\gamma$, over 10 runs. We show the results both for $\gamma$ from 0 to 2.0 with a step of 0.2 and $\epsilon$ from 0.2 to 1.0 with step 0.1, respectively for a complete network (A) a random network (B) and a scale-free network (C)

**Average number of clusters.** To analyze the results of the simulation, we start by taking into account the number of final opinion clusters in the population to understand the degree of fragmentation that the different combinations of the parameters produce. This value indicates how many peaks there are in the final distribution of opinions and provides a first approximation of whether a consensus can be obtained or not. To compute the effective number of clusters, accounting for the presence of major and minor ones, we use the cluster participation ratio, defined in Section 5.1.2, where $c_i$ is the dimension of the $i$th cluster, i.e., the fraction of population we can find

in that cluster. In general, for $m$ clusters, the maximum value of the participation ratio is $m$ and is achieved when all clusters have the same size, while the minimum can be close to 1, if one cluster contains most of the population and a very small fraction is distributed among the other $m$ - 1.

From Figure 6.1, we can see that the behavior of the model across the different network topologies is very similar: the growth of the confidence bound $\epsilon$ allows the population that initially ended up as polarized to reach a full and perfect consensus, at least up to a certain value of the algorithmic bias $\gamma$. The experiments on the three different networks show how the population either converges to one or a few significant clusters or fragments over a wide range of opinions when $\gamma$ is above a certain threshold: the final state shows tens of clusters populated by few agents that cannot merge in the time span allowed in these experiments.

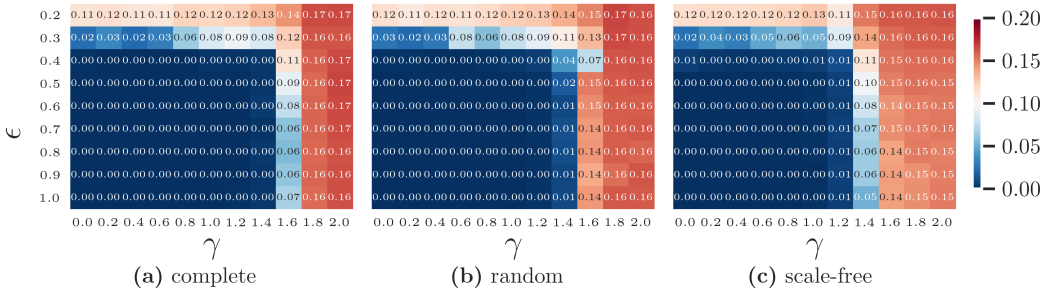
Even in the mean-field for $\gamma \geq 1.6$, the effect of the algorithmic bias is too strong to be mitigated by an increment in the bounded confidence parameter $\epsilon$. The total number of clusters grows with $\gamma$ from values around 10 to values around 70.

However, the population only has $10^5$ iterations to reach convergence, and in some cases, the process reaches this bound without having reached equilibrium, as we will see later in this paragraph. In [204], the maximum number of iterations was set to $10^7$, allowing the population always to reach a steady state. While analyzing what happens when time goes to infinite is important, it is also important to understand how the final status may change with a much shorter dynamic. The present results could mean that consensus - even if theoretically possible - may never be reached in a real setting where there is a finite amount of time to discuss a topic and the population may instead remain fragmented.

Considering only $\gamma \leq 1.4$, we can see that up to this value, the results remain the same described in [204]: for $\epsilon \geq 0.4$ a consensus is always reached, even if it tends to become less and less perfect, while for $\epsilon \leq 0.4$ the number of clusters increases with the bias, which brings the population to a polarization of opinions even in situations where the *DW Model* [54] would have produced a full consensus.

Introducing a different network topology, such as a random network or a scale-free network, however, produces a change in the behavior for very strong biases. Such a result suggests that a sparser topological structure has a small impact on the observed results until the introduced bias is not strong. However, as the algorithmic filtering grows stronger, the sparsity has a very severe impact, preventing consensus - even in cases where it was observed as a possible outcome in mean-field. Moreover, while for $\gamma$ values below the fragmentation threshold, the effective number of clusters is very similar across the three different network topologies, in the fragmented state, we can see that in the scale-free case, the number of clusters is higher - on average - than in the random case and both show overall higher values with respect to the complete networks. It is not clear how this different behavior depends on the topology and how it depends on a different average degree and thus the total number of links in the networks, but we can assume that the more the sparsity, the more it gets difficult for opinion clusters to merge when the bias limits very much the number of agents to interact with.

**Average pairwise distance.** To study the degree of polarization/fragmentation, we computed the average pairwise distance between the agents' opinions. Given an agent $i$ with opinion $x_i$ and an agent $j$ with opinion $x_j$ at the end of the diffusion process, the pairwise distance between the two agents is $d_{ij} = |x_i - x_j|$. The average pairwise distance in the final state can be computed as $\frac{\sum_{i=0}^{N} \frac{(\sum_{j=0}^{N} d_{ij})}{N}}{N}$. In every network,




(a) complete (b) random (c) scale-free

FIGURE 6.2: **Average opinion distance across topologies.** The figure displays the average pairwise opinion distance as a function of $\epsilon$ and $\gamma$, over 10 runs. We show the results both for $\gamma$ from 0 to 2.0 with a step of 0.2 and $\epsilon$ from 0.2 to 1.0 with step 0.1, respectively for a complete network (A) a random network (B) and a scale-free network (C)

the average opinion distance goes from a minimum value of 0.0 when the population reaches a full consensus, and every agent holds the same opinion to a maximum value of $0.15 - 0.17$ when there are tens of opinion clusters in the population. We can observe that such a distance follows the same pattern identified in the number of clusters: it decreases as $\epsilon$ grows and grows with $\gamma$. However, an important thing to point out is that while the difference in the number of clusters can be very high, the opinion distance differences are not so high between a state with three final clusters and a state with 80 final clusters. Indeed, when the opinion distribution is very fragmented, the different clusters tend to get closer to each other. This holds for the three different networks considered in this work.

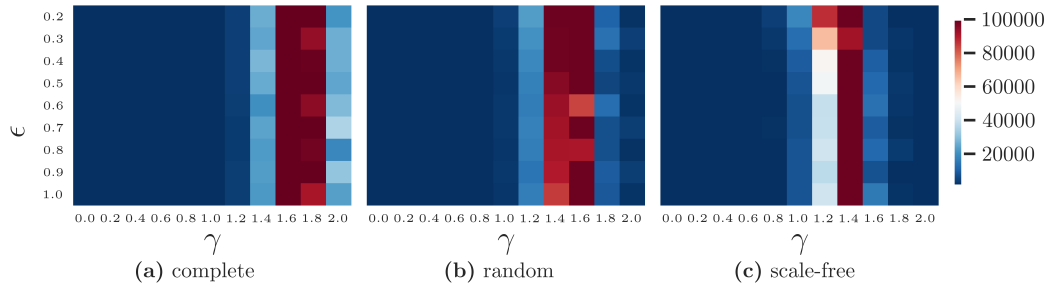
There are also some cases in the complete network where the average pairwise distance decreases despite the number of clusters in the final state is higher. This result suggests that the final peaks in the opinion distribution are indeed all very close to each other, and with a longer simulation, a lower level of fragmentation could be reached.

That considered, we can state that the average pairwise distance is suitable to highlight the transition from consensus to polarization or a limited number of clusters. However, it is not suitable to characterize a growing fragmentation or an intermediate state where there are still many cluster growing closer to each other before merging.

**Time to convergence.** Finally, we consider the time to convergence. The time to convergence is measured as the number of iterations (each constituted by $N$ pairwise interactions).

Figure 6.3 compares the evolution of the time to convergence as a function of $\epsilon$ and $\gamma$. The three plots all show a similar behavior: the main impact on time to convergence (since $\mu$ and $N$ are fixed) is given by $\gamma$. In particular, for every value of the parameter $\epsilon$ in every network the convergence slows down until it reaches its peak for a certain value of the bias, then the time to convergence starts to decrease as the bias grows.

**Mesoscale structure.** We saw from the previous analysis that changing the topology of the network, even with a low average degree and a scale-free structure - doesn't affect much the dynamics. To understand how - instead - the addition of a mesoscale structure may affect the process of opinion diffusion we simulated three




(a) complete (b) random (c) scale-free

FIGURE 6.3: **Average number of iterations to convergence across topologies.** The figure displays the average number of iterations to convergence as a function of $\epsilon$ and $\gamma$, over 10 runs. We show the results both for $\gamma$ from 0 to 2.0 with step of 0.2 and $\epsilon$ from 0.2 to 1.0 with step 0.1 respectively for a complete network (A) a random network (B) and a scale-free network (C)

different scenarios over the previously described LFR networks and we analyzed the same measures as a function of $\gamma$, $\epsilon$ and also $\mu_{LFR}$. We fixed three different settings:

i) the opinions are uniformly distributed across the whole population, like on complete, random, and scale-free cases previously analyzed;

ii) a random mean opinion to each community is assigned, and then the opinions are normally distributed within the community with standard deviation equal 0.01;

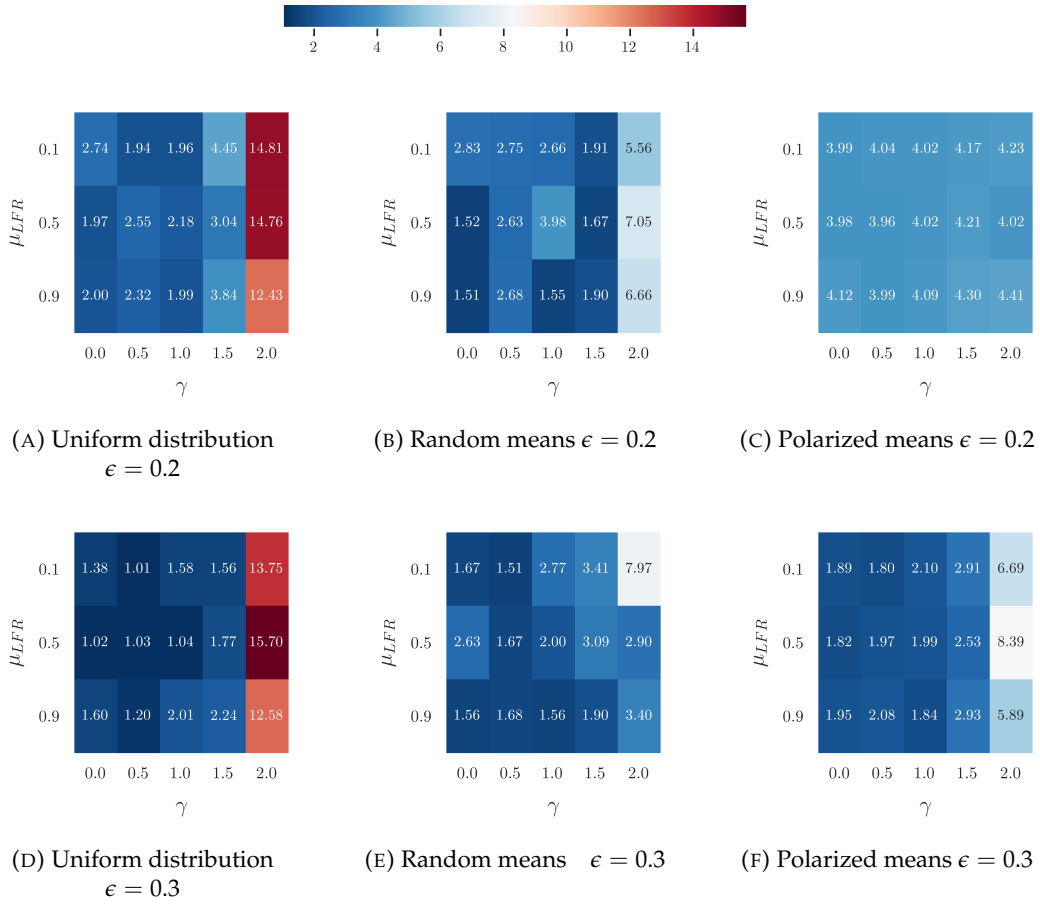
iii) the opinions are normally distributed with predefined means $\in \{0.25, 0.5, 0.75, 1.0\}$ and a standard deviation equal 0.01.

From Figure 6.4 it seems that the dynamics remains qualitatively the same as in the previous cases. A higher $\epsilon$ foster consensus while as $\gamma$ grows so does fragmentation. However, even with opinion randomly distributed across population, it seems that the mesoscale structure reduces the fragmenting effects of the bias (Figure 6.4(A)-(B)) resulting in a lower number of clusters for very high values of bias. If we start assigning to each community a random mean opinion and distributing opinions across community members with a small variance (Figure 6.4(C)-(D)) we can see that fragmentation is overall reduced. When the mean opinions of the communities are more distant than the confidence bound (Figure 6.4(E)-(F)) we always obtain four to five final clusters since different communities cannot merge, and eventually, when the selection bias is very strong, some of them split into more than one cluster. However, in the case of polarized communities, if the mean initial opinions are less distant than the confidence bound, the dynamic remains the same: we see a slow rise from polarization to fragmentation as the bias grows.

## 6.2 Discussion and Conclusions

In this Chapter, the Algorithmic Bias model - developed within the framework of bounded confidence - was simulated on complete, random scale-free network topologies. Such an analysis was conducted to discover and characterize the differences affecting model simulation outcomes while moving from a mean-field scenario (as proposed by the original authors) toward more complex ones.

Algorithmic bias is argued to be an existing factor affecting several (online) social environments. Since interactions occurring among agents embedded in such realities are far from being easily approximated by a mean-field scenario, in our study,




(A) Uniform distribution $\epsilon = 0.2$ (B) Random means $\epsilon = 0.2$ (C) Polarized means $\epsilon = 0.2$

(D) Uniform distribution $\epsilon = 0.3$ (E) Random means $\epsilon = 0.3$ (F) Polarized means $\epsilon = 0.3$

FIGURE 6.4: **Average number of clusters for a given value of $\epsilon$ as a function of $\mu_{LFR}$ and $\gamma$.**

we aimed to understand the role played by alternative network topologies on the outcome of biased opinion dynamic simulations.

From our study, it emerges that the qualitative dynamic of opinions remains substantially in line with what was observed assuming a mean-field context: an increase in the confidence bound $\epsilon$ favors consensus. In contrast, the introduction of the algorithmic bias $\gamma$ hinders it and favors fragmentation. Conversely, both simulations' time to convergence and opinion fragmentation appear to increase as the topology becomes sparser and the hub emerges. Therefore, our analysis underlines that, alongside the algorithmic bias, the network's density heavily affects the degree of consensus reachability, assuming a population of agents with the same initial opinion distribution.

We also investigated how an underlying community structure affects the dynamics. What emerged is that the community structure enhances the consensus, and a larger algorithmic bias has the only effect of slowing down the convergence process. As already stated by the authors in [204], the initial condition is crucial to determine the final state. Our work showed that polarized communities that are further than the confidence bound cannot converge and that an increasing bias may favor splits into two or more clusters within the same community, even when the starting opinions were very close. In Chapter 7, we will see additional extensions of the *AB Model* that cope with more realistic scenarios involving dynamic network topologies and higher-order interactions.

# Chapter 7

# Modeling Algorithmic Bias in Opinion Dynamics: Simplicial Complexes and Evolving Network Topologies

Despite belief-consistent selection and confirmation bias playing a critical role in opinion formation, the diversity of content/sources encountered during daily activities is not the only driver of such complex realities. Peer pressure-like phenomena play a role in shaping people's opinions [8, 103] and therefore should be considered when addressing the study of how public opinion evolves in social networks. People are likely to experience social pressure in both face-to-face and digital interactions. For example, suppose three individuals are mutual friends, and there is a disagreement on a particular topic. In that case, the majority opinion within the group will likely prevail, and the minority will adopt the majority opinion. Within social networks like Twitter, users participate in binary opinion exchanges with other users, i.e., through direct messaging, which can be modeled as binary interactions. However, the possibility of sharing tweets and engaging in public discussions opens the question of how participants can be influenced by others' opinions expressed in the thread and, in turn, influence their peers' opinions.

Understanding how different social mechanisms may influence the direction of public opinion and the levels of polarization in society has always been a crucial task, and a great challenge for computational social scientists [49]. Unfortunately, empirical studies on how opinions form and evolve - influenced by environmental and sociological factors - are still lacking [183]. Indeed, if on one side moving toward data-driven analyses is necessary, on the other, models are essential to comprehend causes and consequences within controlled scenarios. Unfortunately, classic opinion dynamics models are often very simplistic and cannot capture the complexity of the observed phenomenon.

In the last twenty years, indeed, "digital era" specific characteristics are being included in recent opinion dynamics models, i.e., algorithmic personalization [152, 204, 185, 184, 186] to understand what changes this new world brought into the way public opinion is shaped; however, several others are still missing making thus leaving room for more accurate modeling (see Section 4.1 for an extended discussion on the matter). Among such often neglected peculiarities, we can list the temporal dynamic of social interactions. Not only do network topologies evolve, but often this evolution is co-dependent on the dynamic process taking place over the networks, such as opinion exchange [157]. For this reason, recent efforts focus on studying opinion dynamics on dynamic/adaptive networks [194, 130, 111, 120] also in the

context of the Deffuant-Weisbuch and other bounded confidence models [135, 130, 194], describing the effects of the evolution of the underlying network structure on the final state of the population and explaining the effects of the opinion exchange on the structure topology (Section 2.2.3).

Despite group phenomena being present in classical opinion dynamics models (see [83]), it has been recently recognized the importance of using higher-order structures to explain and predict collective behaviors that could not be described otherwise [18, 105] - e.g., peer pressure.

In [204] and [177] (see Chapter 6), it emerged that the dynamics and final state are mainly determined by $\epsilon$ and $\gamma$, with the confidence threshold enhancing consensus and the bias enhancing fragmentation. Comparing simulations performed on complete, ER, and scale-free networks, it emerged that the role of the underlying topology is negligible concerning the effects of the model parameters (thus, confirming what was previously observed in [228, 73, 210] for the scale-free scenario). However, a higher sparsity implies that fragmentation emerges for lower values of the algorithmic bias.

Moving from the results discussed in Chapter 6, where static network models and binary interactions were assumed, in the following Chapter, we aim to study the effects of adaptive networks (Section 2.2.3)- where the dynamic of the network depends on the opinion dynamics - and higher-order interactions (Section 2.1.2) have on opinion formation and evolution when in the presence of a filtering algorithm. To such an extent, we focus our analysis on the same network models employed in Chapter 6, namely Erdős–Rényi [64] and Barabási–Albert[15]. Adopting such controlled environments, used to simulate the social structure among a population of interacting individuals, we analyze the behaviors of the two extensions of the Algorithmic Bias model [204] and discuss the role of arc rewiring towards like-minded individuals and peer pressure within 2-cliques.

This Chapter is organized as follows. In Section 7.1 we introduce the two extensions and describe our experimental workflow; in Section 7.2 we discuss the main finding of our simulations; Section 7.3 concludes the Chapter while opening to future investigations.

## 7.1 Model and Methods

### 7.1.1 Algorithmic Bias: from Fixed Topologies to Adaptive Networks

Despite this being a crucial step towards reality, assuming that social networks are static during the whole period is unrealistic. Interactions and relationships evolve, and this evolution influences and is influenced by the dynamical process of opinion exchanges and the presence of recommender systems and filtering algorithms for the construction of the social network, reinforcing the tendency toward homophilic choices.

In the present work, we extended the baseline model [204] introducing the possibility of arc rewiring, creating the Adaptive Algorithmic Bias model (*AB Model* henceforth), where peer-to-peer interactions are affected by algorithmic biases, and the networks evolve influenced by such interactions, bringing people to connect to peers with opinions within their confidence bound.

**Definition 20 (Adaptive Algorithmic Bias model – *AB Model*)** *Let us assume a population of N agents, where each agents i has a continuous opinions $x_i \in [0, 1]$ and a population of M mass media with fixed continuous opinions $x_{m_i} \in [0, 1]$. At every discrete*

*time step, an agent $i$ is randomly picked from the population, while $j$ is chosen from $i$'s peers according to Definition 18. If their opinion distance is lower than a threshold $\epsilon$ then both of them change their opinion according to Equation* (5.1)*. If their opinion distance is above $\epsilon$:*

- *with probability $p_r$ we remove the link $(i, j)$ and randomly add a link $(i, z)$, iff $|x_i - x_z| < \epsilon$;*
- *with probability $1 - p_r$ the* DW Model *is applied, i.e., both opinions and network structure remain unchanged.*

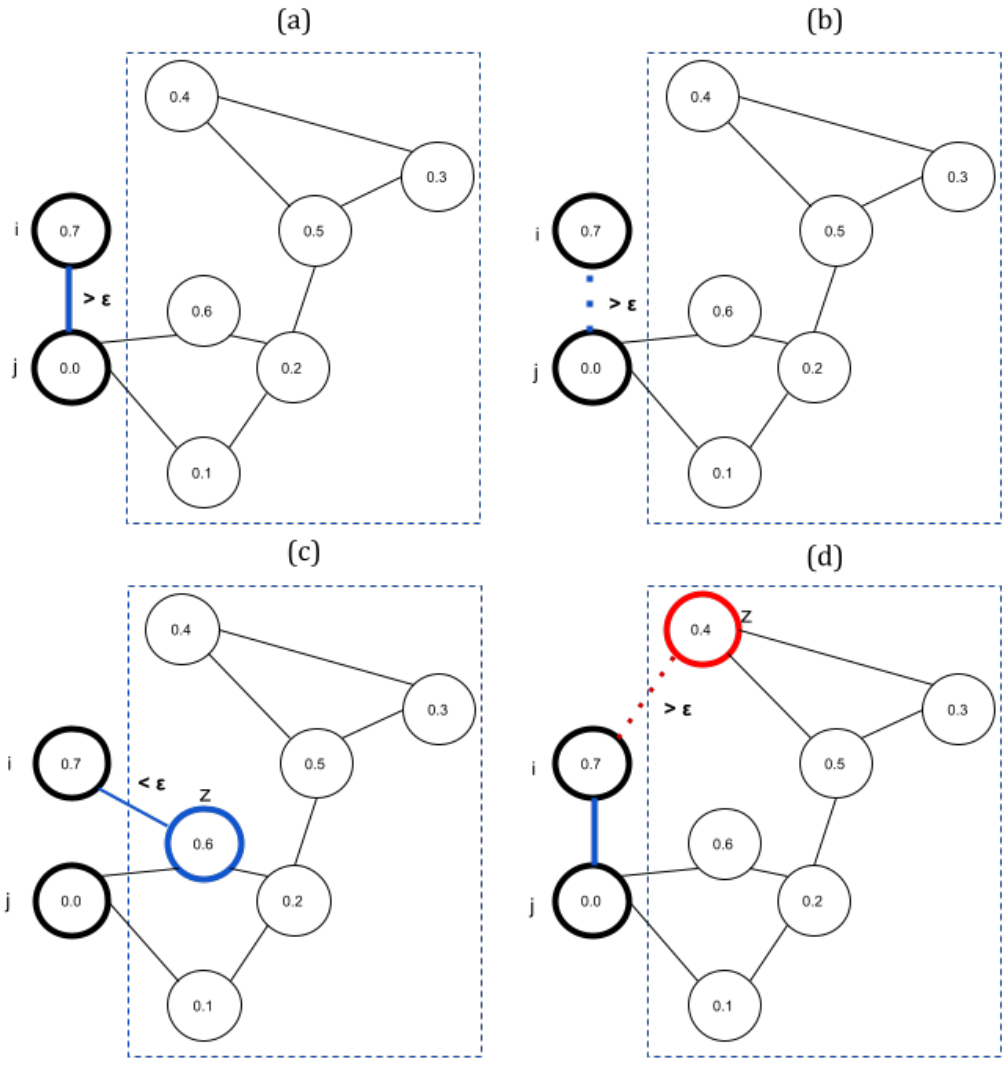



FIGURE 7.1: **A schematic illustration of the rewiring step under bounded confidence.** In this example the confidence bound is $\epsilon = 0.2$. In (A), we can see that the interacting pair $(i, j)$ has an opinion distance further than the confidence bound. For this reason (B) node $i$ tries to break the arc $(i, j)$ and form a new arc $(i, z)$ (with probability $p_r$, with probability $1 - p_r$ nothing happens). Node $z$ is chosen randomly between the remaining nodes in the network. In the case that $|x_i - x_z| < \epsilon$ the arc $(i, j)$ is broken and the arc $(i, z)$ is formed. Otherwise, if $|x_i - x_z| \geq \epsilon$, the rewiring fails, and the network structure remains the same.

We added to the *AB Model* a new parameter, namely $p_r \in [0, 1]$, indicating the probability that the agent in a situation of cognitive dissonance decides to rewire

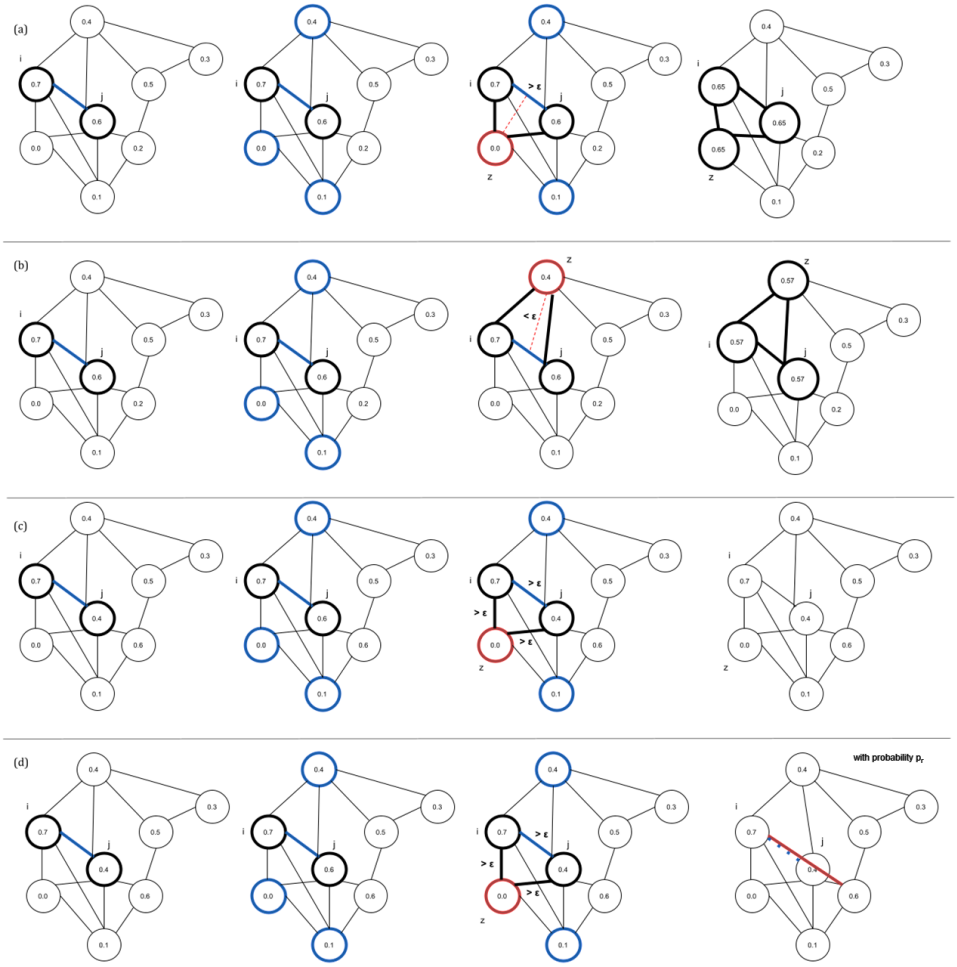



FIGURE 7.2: **Example of the AABSC model.** Examples of different cases in the Adaptive Algorithmic Bias Model on Simplicial Complexes. In (A), a triangle $(i, j, z)$ is chosen, and the minority node adopts the mean opinion of the majority. In (B), there is no minority, so the three agents adopt their average opinion. In (C), there is no majority: nothing happens. In (D), there is no majority, and agent $i$ rewires the discording arc with $j$ towards a more like-minded agent. The process in (D) is the same described in Figure 7.1.

their link instead of just ignoring their peer opinion. To maintain the model as simple as possible, this parameter is assumed to be constant across the population and does not depend on the opinion distance. To rewire the arc, a node $z$ is randomly selected from the set of non-neighboring nodes, and if $|x_i - x_z| < \epsilon$, the agent $z$ links to the agent $i$, otherwise the structure of the graph remains unchanged (see Figure 7.1 for an example of this process and Algorithm 1 for the process pseudocode).

Without considering the algorithmic bias in the choice of the interacting peer, our work is similar to [135, 130]. In [135] the process of rewiring works in the same fashion as in the present work, however, every time the rewiring option is chosen over the standard *DW Model* update rule, the old link $(i, j)$ is broken and a new link $(i, z)$ is formed towards a random non-neighboring agent, even if this agent's opinion is beyond $i$'s confidence threshold. Even in [130] the rewiring process has a different formulation diverging from the proposed one due to the following specificities: (A) at iteration, a set $M$ of discordant edges is rewired and then a set $K$ of edges undergoes the process of opinion update (i.e., if $M < K$ opinion change faster than node rewire, like in the present work); (B) during the rewiring stage the node

**Algorithm 1** Rewiring

Given the pair $(i, j)$ where $d_{ij} \geq \epsilon$
Randomly select a vertex $z$ from the remaining vertices of the graph
**if** $|x_i - x_z| < \epsilon$ **then**
    The arc $(i, j)$ is removed from the graph
    The arc $(i, z)$ is added to the graph
**else**
    The structure of the graph remains unchanged

selection does not happen entirely at random, rather it is "biased" towards similar individuals (still allowing the connection with peers with opinions beyond the confidence threshold); finally (C) the confidence bound and the tolerance threshold for the rewiring are modeled as two independent parameters.

Our implementation assumes a "zero-knowledge" scenario where agents are not aware of the statuses of their peers beforehand: once rejected the algorithmically biased interaction suggestion, if an agent decides to break the tie and search for a new peer to connect with it will not rely (for that task) on other algorithm suggestions. We adopt such modeling since in social contexts (e.g., in online social networks), the status of unknown peers is hardly known by a user - at least before a first attempt at interacting with them. Moreover, not delegating the identification of potential peers to connect with to the "algorithm", we allow users to react to a first non-successful system recommendation independently (i.e., during the same iteration, the user prefers not to trust the algorithm. Instead, he/she makes a blind connection choice). Therefore, rewiring a link toward a like-minded individual is not always feasible given the limits of users' local views.

### 7.1.2 Beyond pairwise interactions: modeling peer pressure

As introduced in Section 2.1.2, different structures can be employed to model higher-order interactions. However, in the specific context of this Chapter, we chose to employ Simplicial Complexes, since the idea is that a triangle of connected agents may experience peer pressure because it constitutes a group of friends, a strong friendship relationship, where in addition to the binary friendships there is a higher-order relationship among these agents. Since a triadic friendship, denoted by a triangle on the social network, does include the binary friendships between each of the three individuals, we propose and analyze an extension of the Algorithmic Bias model to include second-order interactions: the Algorithmic Bias model on Simplicial Complexes (inspired by [113] and adapted to the context of bounded confidence models with continuous opinions). This allows us to incorporate peer pressure in an environment where confirmation and algorithmic biases are still present.

In order to implement peer pressure, i.e., a mechanism for which the majority opinion pressures the individual "minority" one to conform, we first need to define what a majority is in the context of a continuous opinion dynamics model. We choose to consider two nodes "agreeing" if their opinion distance is below the confidence threshold, i.e., $|x_i - x_j| < \epsilon$, similarly to [135].

**Definition 21** *Algorithmic Bias model on Simplicial Complexes* – ABSC Model *Let us assume a population of N agents, where each agent i has a continuous opinions* $x_i \in [0, 1]$ *and a population of M mass media with fixed continuous opinions* $x_{m_i} \in [0, 1]$. *At every discrete time step, an agent i is randomly picked from the population, while j is chosen from*

**Algorithm 2** Adaptive Algorithmic Bias model on Simplicial Complexes

Given two vertex $n1, n2$ linked by an edge $(n1, n2)$
Compute the set $T$ of the triangles including $(n1, n2)$
**if** $T = \emptyset$ **then**
    Baseline rule
**else**
    Choose $t \in T$ with $t = \{(n1, n2), (n1, n3), (n2, n3)\}$ choosing $n3$ according to Definition 18
    **for** any possible permutation (i, j, z) of the nodes in T **do**

        **if** $|x_i(t) - x_j(t)| < \epsilon$ and $|x_z(t) - avg(x_i(t), x_j(t))| < \epsilon$ **then**
            $newOpinion = avg(x_i(t), x_j(t), x_z(t))$
            $x_i(t+1), x_j(t+1), x_z(t+1) = newOpinion$
            return
        **else if** $|x_i(t) - x_j(t)| < \epsilon$ and $|x_z(t) - avg(x_i(t), x_j(t))| \geq \epsilon$ **then**
            $newOpinion = avg(x_i(t), x_j(t))$
            $x_z(t+1), x_i(t+1), x_j(t+1) = newOpinion$
            return
        **else**
            continue
        If none of the three pairs forms a majority, nothing changes

*i's peers according to Definition 18. If the set of triangles T incident on $(i, j)$ is nonempty, the model selects a third node z from T according to Definition 18. Otherwise, the model goes back to the* AB Model *rules. Once the interacting triplet is chosen, if two agents form a majority ($|x_a - x_b| < \epsilon$, two scenarios may arise:*

- *agent c already "agrees" with the majority, i.e., its opinion distance from the average opinion of the majority is below the confidence threshold*
- *the third agent is in a situation of cognitive dissonance with the majority, i.e., its opinion distance from the average opinion of the majority is beyond the confidence threshold*

*In the former scenario, the attractive behavior of the pairwise model is adapted to the triadic case: the agents take the average opinion of the triplet; in the latter, we implemented peer pressure by making the three agents adopt the average opinion of the majority.*

*Rewiring takes place with probability $p_r$ between the disagreeing pair $(i, j)$ with $|x_i - x_j| \geq \epsilon$ when T is an empty set or a "majority" cannot be found in T.*

The *ABSC Model* rules are detailed in Algorithm 2.

Our goal here is to understand the effects of higher-order interactions in a biased environment on the degree of fragmentation reached by the population in the final state. To such an extent, we tested this extended model on the same two graph models: ER [64] and a scale-free [15] network.

### 7.1.3 Experimental settings

Like in [204], to avoid undefined operations in Definition 18, when $d_{ik} = 0$ we use a lower bound $d_\epsilon = 10^{-4}$. The simulations are designed to stop when the population reaches an equilibrium, i.e., the cluster configuration will not change anymore,

even if the agents keep exchanging opinions. We also set an overall maximum number of iterations at $10^5$ as was done in Chapter 6. We compute the average results over 30 independent executions for each configuration to account for the model's stochastic nature. The initial opinion distribution is always drawn from a random uniform probability distribution in $[0, 1]$. To better understand the differences in the final state concerning the different topologies considered, we study the model on all networks for different combinations of the parameters. We are interested in understanding the effects of a co-evolving topology affected by homophily on the dynamics of public opinion in a population and the consequences of peer pressure when moving from pairwise to higher-order interactions.

Moreover, we are also interested in whether – parameters being equal – different initial network topologies influence the final cluster configuration in such extended models. We tested our model, seeding the co-evolution with two different network topologies: an Erdős–Rényi (random) and a Barabási–Albert (scale-free). We set the number of nodes $N = 250$ in both networks. For the ER network, we fix the $p$ parameter (probability to form a link) to 0.1 (thus imposing a *supercritical* regime, as expected from a real-world network); we obtain a network composed of a single giant component with an average degree of 24.94. In the BA network, we set the $m = 5$ (i.e., the parameter regulating the number of edges to attach from a new node to existing nodes), thus obtaining a network instance with an average degree equal to 9.8.[1]

In our simulations, we evaluated the different models on the different possible combinations of the parameters over the following values:

- $\epsilon$ takes a value from 0.2 to 0.4 with a step of 0.1. We chose these values because these are the values for which, in the *AB Model*, we can observe a shift from polarization to fragmentation and from consensus to polarization. Higher values of $\epsilon$ lead to consensus regardless of the strength of the algorithmic bias until the bias is high enough and fragmentation explodes.
- $\gamma$ takes a value from 0 to 1.6 with a step of 0.4; for $\gamma = 0$ the model becomes the *DW Model*. We would see only fragmented final states for higher values of $\gamma$.
- $\mu = 0.5$, so whenever two agents interact, they update their opinions to the pair's average opinion if their opinions are close enough
- $p_r$ (for the Adaptive version of the models) takes a value from 0.0 to 0.5 with a step of 0.1; for $p_r = 0.0$ the model becomes the *AB Model* in the case of the *AB Model*.

To analyze the simulation results, we start by considering the number of final opinion clusters in the population to understand the degree of fragmentation produced by the different combinations of the parameters. This value indicates how many peaks there are in the final distribution of opinions and provides a first approximation of whether a consensus can be obtained or not. To compute the effective number of clusters, accounting for the presence of major and minor ones, we use the cluster participation ratio, as in [204] (see Section 5.1.2) where $c_i$ is the dimension of the $i$th cluster, i.e., the fraction of the population we can find in that cluster. In general, for $m$ clusters, the maximum value of the participation ratio is

[1] Note that the empirical average degree slightly deviates from the expected asymptotic value ($< k >= 2m = 10$) due to statistical fluctuations introduced by the random seed used by the generative process.

$m$ and is achieved when all clusters have the same size, while the minimum can be close to 1, if one cluster contains most of the population and a small fraction is distributed among the other $m-1$. To study the degree of polarization/fragmentation, we computed the average pairwise distance between the agents' opinions. Given an agent $i$ with opinion $x_i$ and an agent $j$ with opinion $x_j$ at the end of the diffusion process, the pairwise distance between the two agents is $d_{ij} = |x_i - x_j|$. The average pairwise distance in the final state can be computed as $\frac{\sum_{i=0}^{N} \frac{(\sum_{j=0}^{N} d_{ij})}{N}}{N}$. While the asymptotic number of opinion clusters and the degree of polarization are essential metrics to describe the results of the dynamics qualitatively, the time to obtain such a final state is equally so. In a realistic setting, available time is finite, so if consensus forms only after a very long time, it may never actually emerge in the population. Thus, we measure the time needed for convergence (to either one or more opinion clusters) in our extended model, recalling that every iteration is made of $N$ interactions, whether pairwise or higher-order (triadic).

## 7.2 Results and discussion

### 7.2.1 Adaptive Algorithmic Bias model: close-mindedness leads to segregation in co-evolving networks

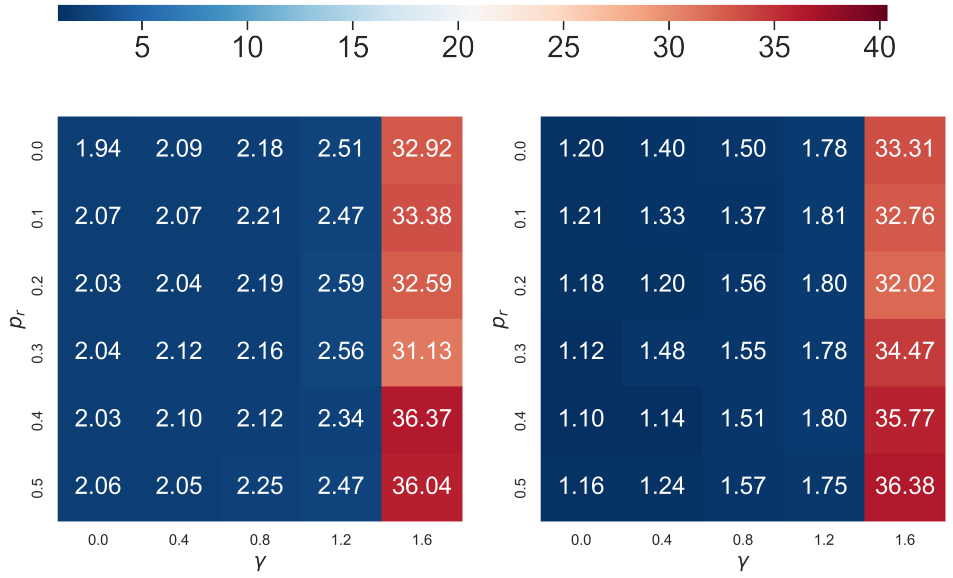
Our simulations suggest that allowing users to break friendships that cause disagreement in a biased online environment has little effect on the levels of polarization/fragmentation when the evolution of the network is remarkably slower than the process of convergence towards a steady state into one or more opinion clusters. However, when two or more opinion clusters form, allowing the rewiring process to continue eventually breaks the network into multiple connected components.

To understand the effects of the interplay of cognitive and algorithmic biases and the probability of link rewiring, we start by looking at the average number of final clusters. Figure 7.3 shows the average number of clusters as a function of $p_r$ and $\gamma$ for $\epsilon \in \{0.2, 0.3\}$.

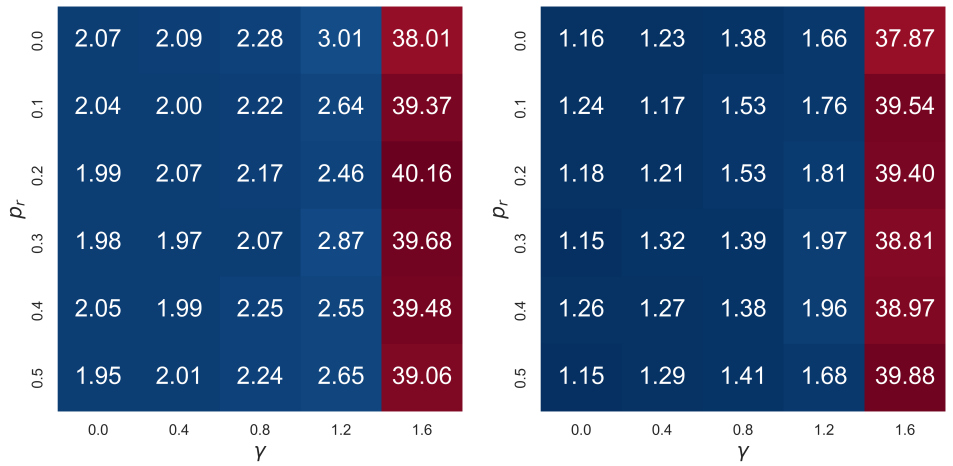
We can observe from the first row of each heatmap that, without rewiring ($p_r = 0.0$), the behavior of the Algorithmic Bias model discussed in Chapter 6 is recovered: fragmentation is enhanced by the bias, while higher values of $\epsilon$ counter the effects of the bias and drive the population towards a consensus around the mean opinion of the spectrum. We already saw in Chapter 6 that concerning the mean-field case, when the topology is sparser even for $\epsilon \geq 0.4$ for a sufficiently large bias, the final states result in a high number of clusters or even not to be clustered at all, i.e., in the final state opinions are still uniformly distributed across the population since the bias is so strong that even like-minded people (whose opinion distance is below the confidence bound) can never converge to each other because they will unlikely interact.

**Erdős–Rényi network.** In Figure 7.3 (A)-(B), we can see that in the *DW Model*, i.e., $\gamma = 0.0$ adding the possibility to rewire arcs during conflicting interactions does not change the final number of clusters on average. For $\epsilon = 0.2$ we obtain a polarized population for every value of $p_r$. A consensus is always reached for $\epsilon \geq 0.3$, specifically for $\epsilon = 0.3$, the main cluster coexists with smaller clusters. In comparison, for $\epsilon = 0.4$, a perfect consensus around the mean opinion is always reached.

The co-evolving topology does not impact the dynamics of the Algorithmic Bias model either: the adaptive topology does not change the fact that the system reaches




(A) ER $\epsilon = 0.2$ (B) ER $\epsilon = 0.3$



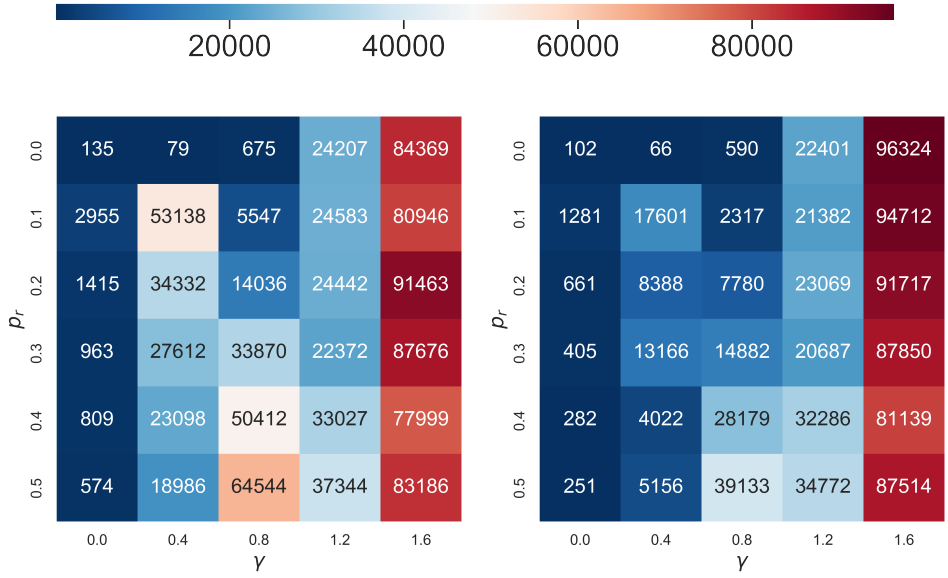

(C) BA $\epsilon = 0.2$ (D) BA $\epsilon = 0.3$

FIGURE 7.3: **Average number of clusters in the steady state of the Adaptive Algorithmic Bias model.** The average number of clusters in the final state of the Adaptive Algorithmic Bias model as a function of $\gamma$ and $p_r$ for (A) $\epsilon = 0.2$ and (B) $\epsilon = 0.3$ starting from the Erdős–Rényi graph and (C)-(D) starting from the scale-free Barabási–Albert graph. These values are averaged over 30 runs.

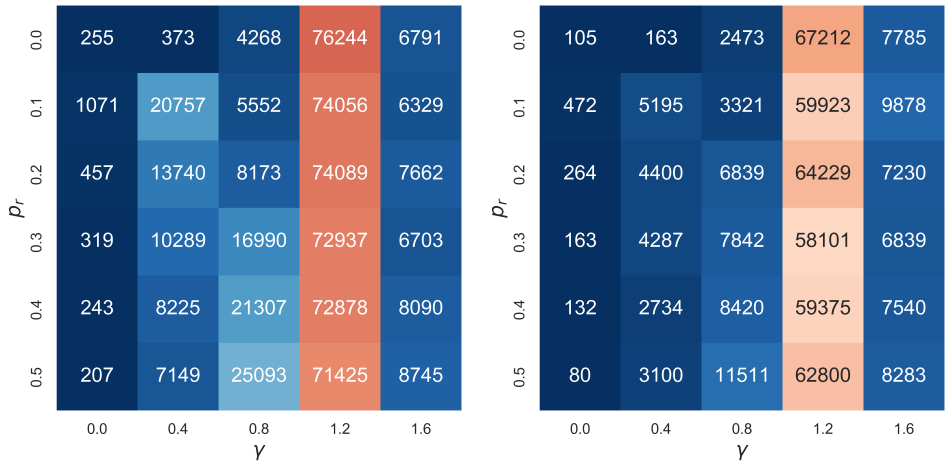
consensus or polarization. Plots in Figure 7.3(A)-(B) show the population moving from two to three clusters in $\epsilon = 0.2$ and from one to two clusters in $\epsilon = 0.3$ and always reaching a consensus for $\epsilon = 0.4$, even if we consider $p_r = 0.5$.

For what concerns the time (i.e., the number of total interactions) the population needs to reach an equilibrium, we can see how the general behavior of the baseline model is kept, even when the network co-evolves with a biased opinion dynamic. Convergence is relatively fast when interactions are not biased, while it slows down as the bias grows until it reaches a peak, after which it speeds up again. Until the peak, a higher number of iterations positively correlates with a higher number of clusters. In contrast, even if convergence is faster, the population is spread across the opinion space after the peak. Since the bias is strong, two agents cannot get closer

in the opinion distribution after the first few interactions, and the condition to reach the steady state is met very quickly. Moreover, every node has a limited network of agents to interact with; with a strong bias, they always exchange opinions with the same agents. Not much can change once they adopt their average opinion, even in the long term.




(A) ER $\epsilon = 0.2$ (B) ER $\epsilon = 0.3$




(C) BA $\epsilon = 0.2$ (D) BA $\epsilon = 0.3$

FIGURE 7.4: **Average number of iterations to convergence in the Adaptive Algorithmic Bias model.** Average number of iterations to convergence in the Adaptive Algorithmic Bias model as a function of $\gamma$ and $p_r$ for (A) $\epsilon = 0.2$, (B) $\epsilon = 0.3$ and (C)-(D) in a scale-free Barabási–Albert graph. These values are averaged over 30 runs.

We can see from Figure 7.4(A)-(B) that this measure is less dependent on the population's open-mindedness as the bias level mainly influences it. However, we can see from the average values that an increase in $\epsilon$ often means a small convergence speed-up, all other parameters being equal.

When we introduce the possibility of rewiring, convergence is generally slower.

Deleting edges beyond one's confidence bound denies agents the possibility of participating in a possible path toward convergence. This does not mean that the population cannot converge; instead, a higher number of total interactions is needed. Without bias, it is worth noticing that a small probability of arc rewiring $p_r = 0.1$ in a close-minded population ($\epsilon = 0.2$) has the slowest time to convergence. Arc rewiring towards like-minded individuals and selection bias combined slow down convergence, especially in close-minded populations (i.e., $\epsilon = 0.2$): we can see that even for a relatively small bias and a relatively small probability of arc rewiring, the steady state needs tens of thousands of iterations, while without arc rewiring less than one hundred would be enough.

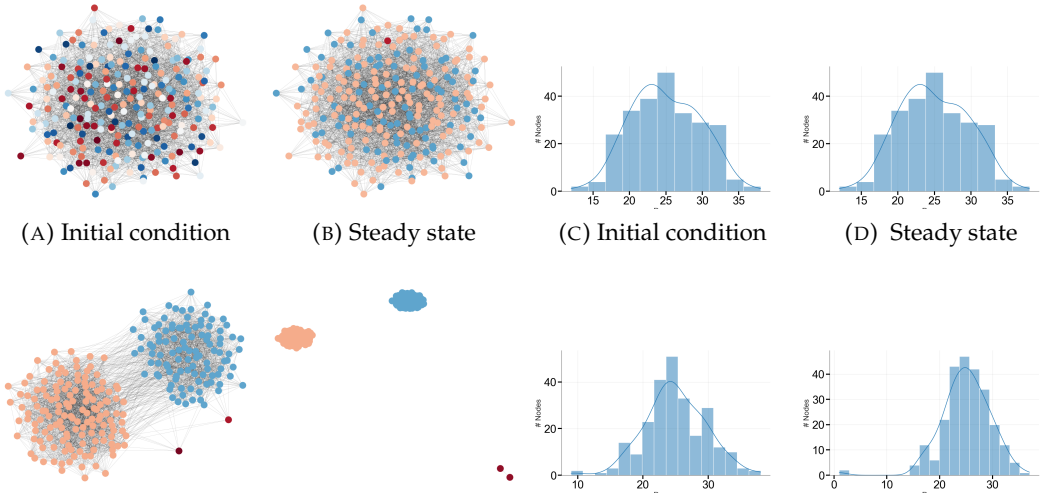
**Barabási–Albert network.** In Figure 7.3(C)-(D), we can see that the same results that were drawn for the ER network also hold for scale-free networks, though, on the latter, fragmentation is higher on average, and the same level of fragmentation arises for lower values of the bias.

Also, for the time at convergence, similar conclusions can be drawn. However, in the scale-free network, the peak is always reached for $\gamma = 1.2$, regardless of the value of $p_r$. As $p_r$ grows, the convergence slows down so much that the system can no longer reach a steady state, and for higher values of the bias, convergence is much faster, while in the ER network, it is overall slower.

To sum up, we can say that the process of co-evolution of the network, along with the diffusion of opinions in the population, does not affect the final opinion distribution in terms of the number of opinion clusters. This is because when there is no bias - or the bias is low - despite a lot of conflicting interactions happening, the process of convergence is too fast with respect to the process of link rewiring to separate the network into many different opinion clusters, enhancing fragmentation; when the bias is high - instead - despite this already has a fragmenting effect, it also reduces the number of conflicting interactions and therefore slows down the process of network evolution, even more, leaving the network structure practically unchanged when the population reaches its steady state.

Despite not changing the final number of opinion clusters, it impacts the network's topology, as we can observe from the examples in Figure 7.5 and Figure 7.6. In this case, we performed experiments with the same initial configuration of the Erdős–Rényi network (i.e., 250 nodes, p = 0.1, and uniformly distributed initial opinions). We stopped the simulations when no opinion change (nor arc rewiring) happened for 1000 consecutive iterations. In these case we set $\epsilon = 0.2$ and we compared results for $\gamma \in \{0.0, 0.5\}$ and $p_r \in \{0.0, 0.5\}$. As we can observe, polarization occurs when the network is static, meaning that there are two opinion clusters in the population while the network structure remains unchanged.

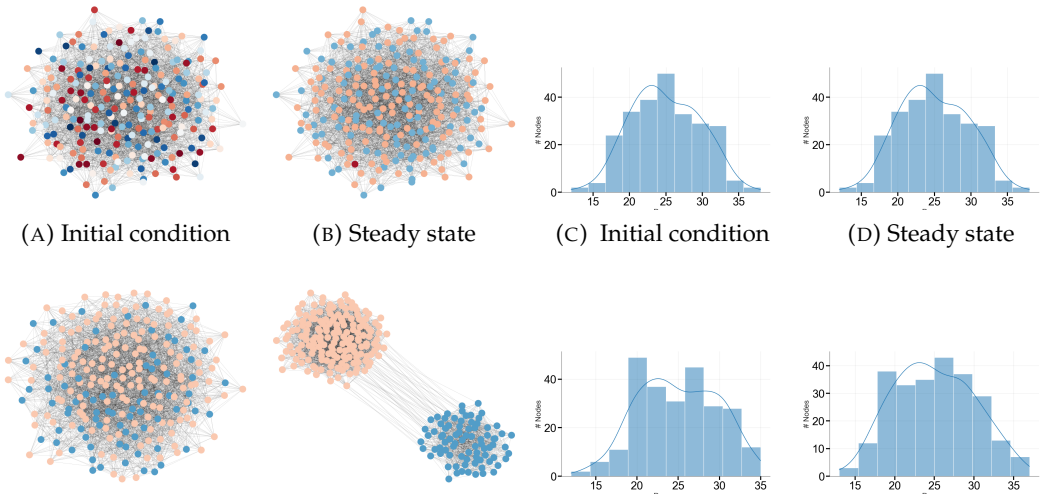
As we can observe from Figure 7.5(E)-(F), the two opinion clusters tend to separate into two different components or - at least - into two different communities on the network, with fewer and fewer inter-communities links when rewiring is allowed. After 100 iterations (Figure 7.5(E)), there is still one connected component, but two polarized communities started to form . It is also worth noticing that, in the steady state (Figure 7.5(F)), every node is connected only to agents holding identical opinions since there are three separate components, each holding perfect consensus. The long left tail of the degree distribution in Figure 7.5(H) is due to the two-nodes component. If we do not consider that component, the final degree distribution is substantially similar to the distribution in Figure 7.5(H) with a slightly lower variance. Introducing an algorithmic bias in the process slows down the convergence, as




(A) Initial condition (B) Steady state (C) Initial condition (D) Steady state

(E) t = 100 (F) Steady state (G) t = 100 (H) Steady state

FIGURE 7.5: **Example of the effects of the adaptive topology on the Adaptive Algorithmic Bias Model on the Erdős–Rényi graph with $\gamma = 0.0$.** An example of the effects of the co-evolution of network structure and opinions in the Adaptive Algorithmic Bias model on the Erdős–Rényi graph for $\epsilon = 0.2$, $p_r \in \{0.0, 0.5\}$ and $\gamma = 0.0$.




(A) Initial condition (B) Steady state (C) Initial condition (D) Steady state

(E) t = 100 (F) Steady state (G) t = 100 (H) Steady state

FIGURE 7.6: **Example of the effects of the adaptive topology on the Adaptive Algorithmic Bias Model on the Erdős–Rényi graph with $\gamma = 0.5$.** An example of the effects of the co-evolution of network structure and opinions in the Adaptive Algorithmic Bias model on the Erdős–Rényi graph for $\epsilon = 0.2$, $p_r \in \{0.0, 0.5\}$ and $\gamma = 0.5$.

we can see from the example in fig Figure 7.6(E) where - starting from the same initial configuration - the network does not present node clusters holding similar opinions after 100 iterations (while this was the case in the absence of bias). This is because bias skews interactions towards more like-minded nodes, further slowing down the process of arc rewiring by reducing the amount of discording encounters. We can see from Figure 7.6(F) that, in this case, equilibrium is reached before two components could form on the network, but there are two well-separated communities, each holding a separate opinion. While a steady state in terms of opinion clusters

may be reached within a few iterations, which are not enough to separate the network into different components, if we allow the process to go on until there are no possible rewirings - since every node is connected to agreeing nodes - the network eventually splits into two or three components when opinions are clustered. Even when the maximum number of iterations set in such simulations is not enough, we can still see that links between polarized opinion clusters are fewer and fewer over time.

### 7.2.2 Adaptive Algorithmic Bias model on Simplicial Complexes: Peer Pressure Enhances Consensus.

In the Adaptive Algorithmic Bias model on Simplicial Complexes, we introduced a simple form of higher-order interaction where three agents can influence each other - as a group - if they form a complete subgraph. Introducing higher-order interactions lets us model the phenomenon of peer pressure, where the majority edge pushes the minority node to conform to their ideology. If there is no minority opinion, we assume there would be an attractive dynamic similar to the one present in the binary case, i.e., the three nodes attract each other and adopt the mean opinion of the group.

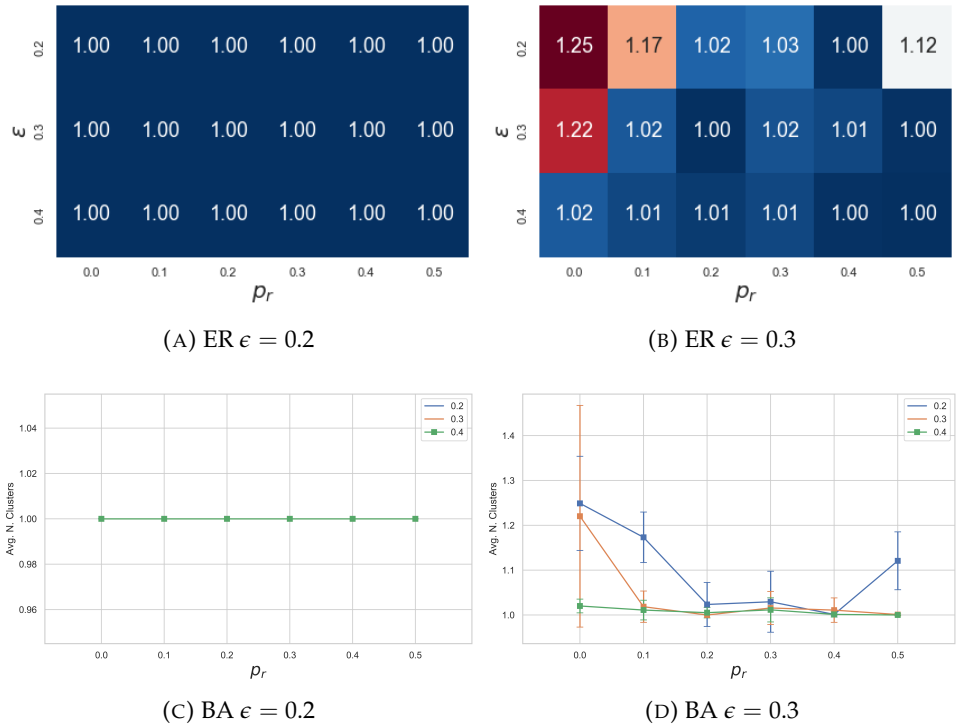
The main result from our simulations is that peer pressure promotes consensus and reduces fragmentation with respect to the binary counterpart. Besides this general conclusion, we can observe in Figure 7.8 that the model's behavior is different in the two chosen networks and that $\gamma$ and $p_r$ still play a role in shaping the final state of the population.

**Adaptive *DW Model* on Simplicial Complexes on complex topologies.** Before analyzing the effects of peer pressure and algorithmic biases on the Algorithmic Bias model, we briefly analyze the results for the *DW Model*, i.e., $\gamma = 0.0$.

We can observe in Figure 7.7(A) that in the ER network, a perfect consensus is always reached, regardless of the level of bounded confidence and rewiring probability (the number of clusters is 1 in every execution of the model and the standard deviation of the final distribution is always 0).

Also, in the scale-free network (Figure 7.7(B)), the consensus is always reached, but it is not always perfect and depends on both the confidence bound $\epsilon$ and the probability of rewiring $p_r$. In a static network ($p_r = 0.0$), introducing peer pressure reduces fragmentation: in the case of $\epsilon = 0.2$, for example, the baseline model would lead to polarization, on average. Changing the update rule to account for group interactions and social pressure reduces the level of fragmentation in the final state. It leads almost the whole population to converge on a common opinion. Not surprisingly, increasing the confidence bound enhances consensus in the same way as in the baseline model. However, in this case, increasing the probability of rewiring reduces fragmentation leading the population to a perfect consensus. In particular, we can see from Figure 7.7 (B) that for $\epsilon = 0.2$ perfect consensus is reached for $p_r > 0.2$, for $p_r > 0.0$ in the case of $\epsilon = 0.3$ and always in the case of $\epsilon = 0.4$.

**Erdős–Rényi network.** Simulating the Algorithmic Bias model on Simplicial Complexes on the chosen ER graph, for $\epsilon = 0.2$ we can see how the population always reaches a consensus for low values of the bias, while for $\gamma \geq 1.2$, peer pressure is not enough to stop the population from polarizing into two opposing clusters. However, if compared with the same results with only pairwise interactions (Figure 7.3), it is clear that fragmentation is strongly reduced.




(A) ER $\epsilon = 0.2$ (B) ER $\epsilon = 0.3$

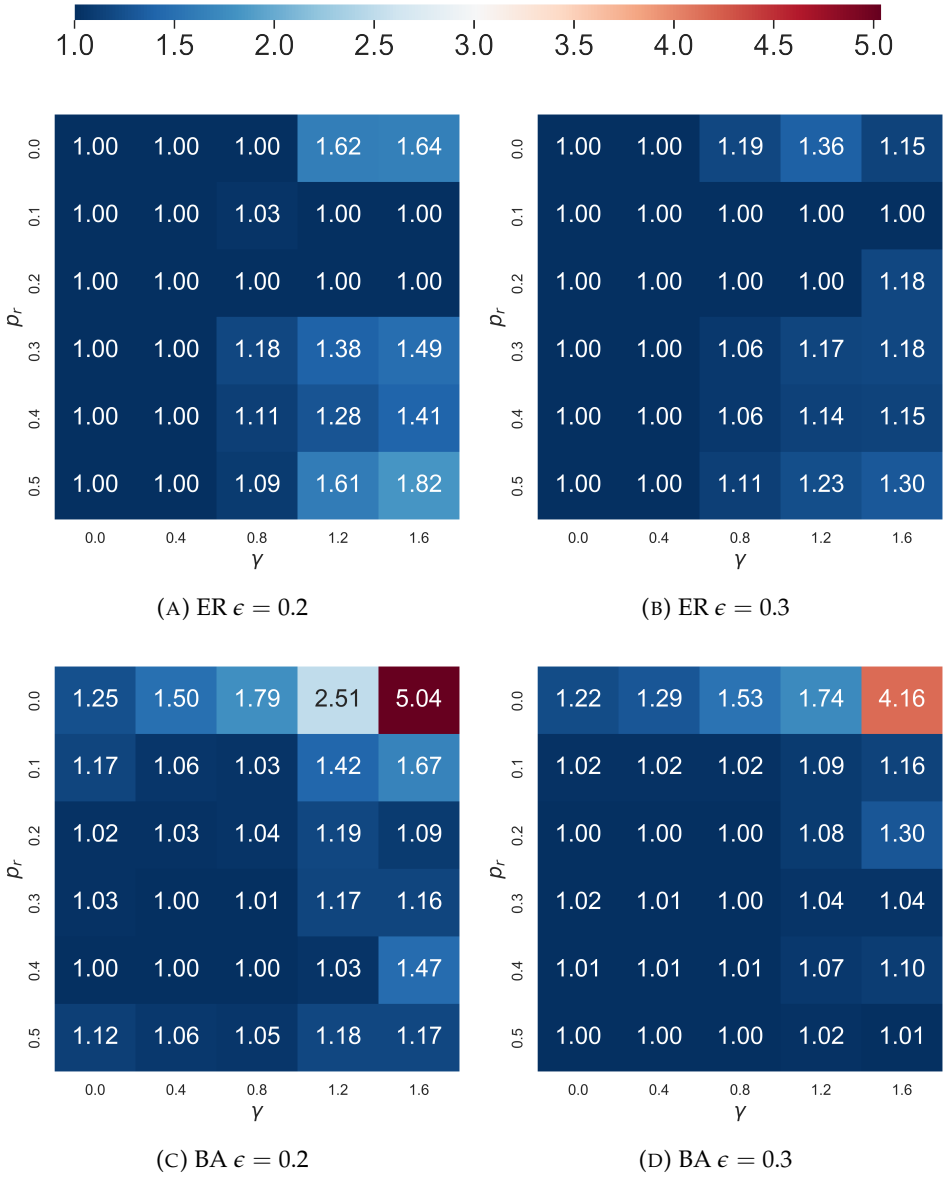
(C) BA $\epsilon = 0.2$ (D) BA $\epsilon = 0.3$

FIGURE 7.7: **Average number of clusters in the steady state for the Adaptive *DW Model* on Simplicial Complexes.** The average number of clusters in the final state for the Adaptive *DW Model* on Simplicial Complexes fixing $\gamma = 0.0$, as a function of $\epsilon$ and $p_r$ for (A)-(B) an Erdős–Rényi graph and (C)-(D) a scale-free Barabási–Albert graph. These values are averaged over 30 runs.

For $\epsilon = 0.3$, the qualitative dynamic remains the same. However, the average number of clusters is reduced overall due to the population's higher open-mindedness, and the population splits into two clusters only in a few cases. In contrast, in most simulations, a majority cluster forms along with a few smaller ones. When the population is open-minded, i.e., $\epsilon \geq 0.4$, consensus is always reached around the mean opinion (i.e., 0.5). The only effect of a higher algorithmic bias is that a few agents cannot converge into the main cluster. Introducing the possibility of rewiring towards a more like-minded individual after a conflicting interaction enhances polarization when combined with a mild or high selection bias (i.e., $\gamma > 0.8$). The population converges into two or three clusters when the confidence threshold is low (either two polarized clusters or two polarized clusters and a moderate one). When the population is mildly open-minded ($\epsilon = 0.3$), the system converges into one or two clusters (either two polarized clusters or a moderate cluster). At the same time, it always reaches a consensus for higher values of the confidence bound (around the mean opinion).

As we did for the previous model, we also analyzed the average time to convergence.

From Figure 7.9, we can see that a higher bias slows down convergence like in the binary model. It is slowed down so much that the population cannot reach a steady state within the imposed time interval. While in the model by Sirbu et al. [204], the level of open-mindedness did not play a crucial role in the time at convergence, in this case, we can see that increasing the open-mindedness of the population also means a faster convergence towards an equilibrium.

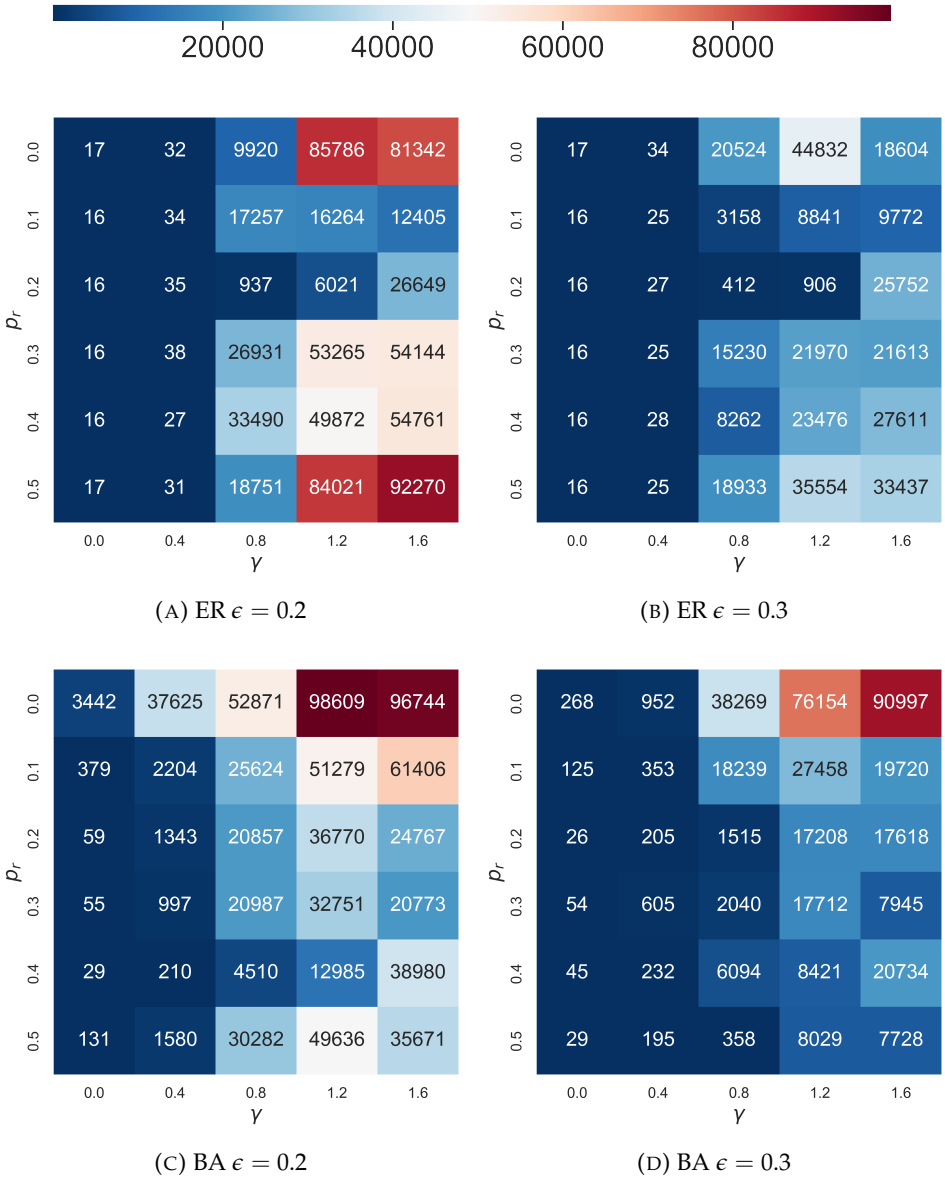



(A) ER $\epsilon = 0.2$ (B) ER $\epsilon = 0.3$

(C) BA $\epsilon = 0.2$ (D) BA $\epsilon = 0.3$

FIGURE 7.8: **Average number of clusters in the steady state for the Adaptive Algorithmic Bias model on Simplicial Complexes.** Average number of clusters in the final state for the Adaptive Algorithmic Bias model on Simplicial Complexes as a function of $\gamma$ and $p_r$ for (A) $\epsilon = 0.2$, (B) $\epsilon = 0.3$ and (C)-(D) in a scale-free Barabási–Albert graph. These values are averaged over 30 runs.

**Barabási–Albert network.** In the scale-free network, the model's behavior is slightly different: a higher probability of arc rewiring seems to reinforce consensus: we can see that when $p_r = 0.0$, the number of clusters in the final opinion distribution is higher as $\gamma$ grows. This small fragmentation is reduced as $p_r$ grows. For example, in the case of a close-minded population, i.e., $\epsilon = 0.2$, we can see that, without rewiring, a consensus is possible until the algorithmic bias is not very strong. However, it is not a perfect consensus (like in the ER network), but there is a major cluster coexisting with many agents scattered across the opinion space. Moreover, such a cluster does not necessarily form around the mean opinion but can be pretty extreme (with the final consensus below 0.2 or above 0.7). For $\gamma = 1.2$, the population becomes polarized: two homogeneous and opposed clusters form, and,

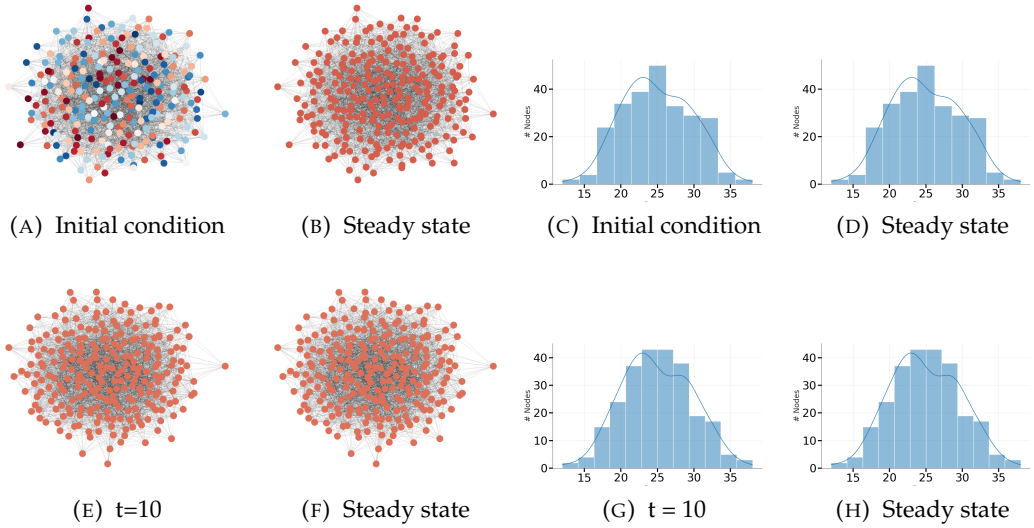



(A) ER $\epsilon = 0.2$ (B) ER $\epsilon = 0.3$

(C) BA $\epsilon = 0.2$ (D) BA $\epsilon = 0.3$

FIGURE 7.9: **Average number of iterations at convergence for the Adaptive Algorithmic Bias model on Simplicial Complexes.** The average number of iterations at convergence for the Adaptive Algorithmic Bias model on Simplicial Complexes as a function of $\gamma$ and $p_r$ for (A) $\epsilon = 0.2$ and (B) $\epsilon = 0.3$ in an Erdős–Rényi graph and (C)-(D) in a scale-free Barabási–Albert graph. These values are averaged over 30 runs.

in some cases, there are few "outlier" agents around the mean opinion or further at the extremes. Finally, for $\gamma = 1.6$, the population splits into multiple clusters: still, in most cases, two major polarized clusters form alongside a variety of minor clusters below, between, and above the two. Two cohesive groups coexist with a population of individuals scattered across the opinion space so that the final distribution is not so different from the initial one: multiple opinions are still present in the population and cover the whole range [0,1]. Raising $p_r$ to 0.1 prevents fragmentation, but for strong biases, the population polarizes. For $p_r \geq 0.2$, consensus is always reached. However, as in the baseline case (without rewiring), consensus does not necessarily form around the center of the opinion space but can vary and form on strongly extreme opinions. An increase in open-mindedness also counters the fragmenting




(A) Initial condition (B) Steady state (C) Initial condition (D) Steady state

(E) t=10 (F) Steady state (G) t = 10 (H) Steady state

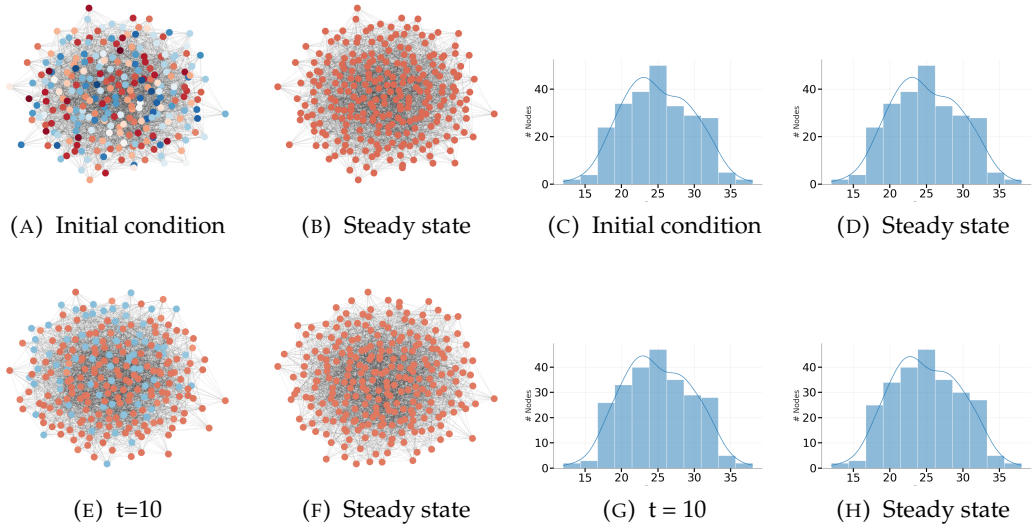
FIGURE 7.10: **Example of the effects of the adaptive topology on the Algorithmic Bias Model on Simplicial Complexes on the Erdős–Rényi graph.** An example of the effects of the adaptive topology on the Algorithmic Bias Model on Simplicial Complexes on the Erdős–Rényi graph for $\epsilon = 0.2$, $p_r \in \{0.0, 0.5\}$ and $\gamma = 0.0$. The convergence towards consensus is faster and is always reached before the network can cluster around different opinions.

effects of the algorithmic bias. The average number of clusters reduces as $\epsilon$ grows, all other parameters being equal. In the case of a highly mildly open-minded population, i.e., $\epsilon = 0.4$, consensus can be prevented only with an extreme algorithmic bias ($\gamma = 1.6$) and without the possibility of arc rewiring. Moreover, in the scale-free topology, convergence is faster with respect to the ER network.

As we can see from Figure 7.10 and Figure 7.11 since, in this case, the consensus is enhanced by peer pressure and triadic interactions that also fasten convergence, opinions reach a steady state before the topology of the network can impact the process. We can observe that neither opinions nor nodes segregate during the process. Figure 7.10 shows that in the absence of an algorithmic bias, consensus can be reached within a few iterations, even with low confidence bound. Figure 7.11 shows how introducing an algorithmic bias does not prevent the population from reaching a consensus but slows down the process, even with the help of peer pressure mechanisms. Comparing $\gamma = 0.0$ and $\gamma = 0.5$ after ten iterations, we can see how, in the first case, the population has already reached a consensus, while in the second, two opinion clusters are still present in the network. Due to the fast convergence process towards consensus, even if rewiring is allowed, it does not significantly impact the network structure, as shown in Figure 7.10(E)-(H) and Figure 7.11(E)-(H).

## 7.3 Discussion and Conclusions

Algorithmic bias is an existing factor affecting several (online) social environments. Since interactions occurring among agents embedded in such realities are far from being easily approximated by a mean-field scenario, in our study, we aimed to understand the role played by alternative network topologies on the outcome of biased opinion dynamic simulations. From our study in Chapter 6, it emerged that




(A) Initial condition (B) Steady state (C) Initial condition (D) Steady state

(E) t=10 (F) Steady state (G) t = 10 (H) Steady state

FIGURE 7.11: **Example of the effects of the adaptive topology on the Algorithmic Bias Model on Simplicial Complexes on the Erdős–Rényi graph.** An example of the effects of the adaptive topology on the Algorithmic Bias Model on Simplicial Complexes on the Erdős–Rényi graph for $\epsilon = 0.2$, $p_r \in \{0.0, 0.5\}$ and $\gamma = 0.5$. Bias slightly slows down the convergence process.

the qualitative dynamic of opinions remains substantially in line with what was observed assuming a mean-field context: an increase in the confidence bound $\epsilon$ favors consensus. In contrast, introducing the algorithmic bias, $\gamma$ hinders it and favors fragmentation. Conversely, both simulations' time to convergence and opinion fragmentation appear to increase as the topology becomes sparser and the hub emerges. Therefore, our analysis underlines that, alongside the algorithmic bias, the network's density heavily affects the degree of consensus reachability, assuming a population of agents with the same initial opinion distribution. The present work extends the work in Chapter 6, proposing two extensions of the model and analyzing such extensions on the same two complex networks as in Chapter 6, leaving out the complete network. The first extension considers a straightforward mechanism of arc rewiring so that the underlying structure co-evolves with the opinion dynamics, generating the Adaptive Algorithmic Bias model. The second adds a peer pressure mechanism, considering triangles as simplicial complexes, where a majority - if there is one - can attract a disagreeing node, pushing them to conform. We found that - in general - the role of bounded confidence and algorithmic bias remains the same as in the baseline models, with the former enhancing consensus while the latter enhancing fragmentation. Going from a static to an underlying adaptive topology does not strongly affect the dynamics, leading to the same number of opinion clusters in the steady state. However, suppose we allow the agents to continue interacting with each other. In that case, opinion clusters eventually lead to the formation of mesoscale structures, then finally separating the network into different connected components. On the other hand, peer pressure enhances consensus, reducing the effects of low bounded confidence and high algorithmic biases. Such a model suggests how different sociological and topological factors interact with each other, thus leading populations towards polarization and echo chamber phenomena, contributing to the creation and maintenance of inequalities on social networks. These models can also be employed to study different phenomena besides opinion diffusion, such as the effects

of peer pressure on the adoption of different behaviors where social network structure and psychological factors play a role. The present work presents some of the limitations already considered in [204] while overcoming others. The existence of bounded confidence, for example, and the fact that it is constant across the population is an assumption that should be empirically validated, along with the role of algorithmic filtering in influencing the path toward polarization/fragmentation. We went beyond the concept of static networks considering an adaptive topology; however, to further investigate the role of arc rewiring, a more thorough analysis of the model's parameters' effects on the network's topology should be made. The present implementation of a rewiring mechanism is just one way to incorporate the fact that users hardly know their neighbors' state before interacting with them; however, a mechanism considering only the set of agents with opinions within the confidence threshold would be a useful comparison to the present model. Moreover, to better understand the role of homophily in the sense of friendship formation and its relation to the online social network environment, the role of the recommender system - and therefore algorithmic bias - a biased mechanism simulating "link recommendations" could be implemented - as in [130]. Finally, the importance of social interactions in opinion formation is undeniable. However, external media can be essential in polarizing opinions or driving the population toward consensus. For this reason, we feel their role needs to be further investigated while embedded in an algorithmically biased environment.

Despite having added several layers of realism to the classical *DW Model*, all the works presented in Chapters 5 to 7 lack empirical validation and are not validated on real data. In Part III, we will see how models of opinion dynamics can be exploited to understand real online discussions better.

# Part III

# Applying models to data: hybrid approaches to analyze Polluted Information Environments

# Chapter 8

# Open-mindedness in Polluted Information Environments: Feedback Loop between Models and Data

This thesis began by discussing PIEs and the biases that create distortions in the process of opinion evolution, accentuated by the presence of online social networks.

One of the most debated and analyzed phenomena on online social networks is the tendency to observe political polarization [48, 162, 12, 163], i.e., the divergence of political attitudes to ideological extremes not aiming at reaching any form of synthesis. Indeed, as recalled in Chapter 4, social media blurred the boundaries of communication and democratized content diffusion. However, the fact that we can potentially engage with a lot of different information does not mean that we can engage with and be positively influenced by opposing or even mildly different stances. Controlled experiments on Twitter, for example, show how exposure to opposing views may, actually, polarize users more [11].

As introduced in Chapter 4, in the process of shaping their belief systems, humans are not entirely rational entities and are influenced by a range of cognitive biases [170, 138]. Notably, confirmation bias is a prevalent issue, which is the inclination to dismiss information that contradicts pre-existing beliefs. This is often manifested in two ways: a) individuals tend to engage with those who share similar views [157], and b) they tend to disregard interactions with those who hold opposing views [54]. Regrettably, this tendency to connect with those who share similar beliefs and to overlook differing perspectives can lead to a compromise in open-mindedness [223]. This can further intensify and polarize their views, ultimately contributing to the creation and perpetuation of echo chambers [214].

One of the most exploited approaches for understanding the effects of different kinds of biases on public opinions - especially political opinions - is through mathematical models of opinion formation [205], where parameters incorporate psychological factors (e.g., cognitive biases) affecting individual opinion evolution (see Chapter 3 for a general description of opinion dynamics models and a discussion of the main milestones in the field). Despite having some advantages, opinion dynamics models lack empirical validation [183]. However, thanks to the advent of the Internet - and with the rise of social media - an increasing part of human interactions leave a massive digital footprint that can be exploited to study the dynamics of opinion formation and diffusion. Following such reasoning, since different models or different parameter values can predict different, even opposite, effects of biases

on opinions [152], there is a crucial need for empirical works to study and quantify socio-psychological and external drivers of the dynamics. We saw in Section 4.2 a brief overview of the main approaches in bridging the gap between models and data.

Recognizing the need for empirical investigation [183], our focus shifted towards examining how cognitive biases impact opinion evolution within a real social network.

Drawing from the Deffuant-Weisbuch model's [54] premise of bounded confidence during social influence, we aimed to construct a methodology for quantifying this threshold using real-world data. We developed two distinct methodologies (see ...) in order to deal with both network and hypergraph frameworks.

The primary focus of this investigation centers on the estimation of open-mindedness distributions in different Reddit discussions.

We begin this Chapter by detailing the analysis undertaken in Section 8.1, wherein we quantified the monthly shifts in open-mindedness over an 18-month span within the r/politics subreddit in the discussions during the Trump presidency.

In the subsequent Section, Section 8.2, we applied both methodologies to estimate the six-monthly levels of open-mindedness across three distinct controversial topics (Minorities Discrimination, Gun Control Regulations, and Politics) within the realm of Reddit. Notably, we investigate the contrasts between the utilization of networks and hypergraphs as structural models for discussions.

Furthermore, our investigation extends to Section 8.3, wherein we employ the developed methodology to estimate user-level confidence bounds within a dual-snapshot network. This estimation subsequently serves as the foundation for a novel approach: simulating a heterogeneous bounded confidence version of the *ABMM Model* introduced in Chapter 5, leveraging the estimated values to discern a model setting that best accounts for observed patterns in real-world data.

This Part provides valuable insights into the dynamics of online discussions and opinion polarization by leveraging novel methodologies to estimate open-mindedness distributions across diverse contexts, shedding new light on the structural and temporal factors that drive opinion change and consensus formation in online communities.

The content of this Part refers to 2 articles – [178] and [179] – and a Master's Thesis [69].

## 8.1 Open-mindedness in political discussions during Trump's presidency on Reddit

Political polarization, the divergence of political attitudes to ideological extremes, is a widely debated and analyzed phenomenon on online social networks. Social media has blurred the boundaries of communication and democratized content diffusion, but this doesn't necessarily mean that we engage with and are influenced by opposing or mildly different stances. Humans, who are not perfectly rational, are affected by a series of cognitive biases in the process of forming their belief systems. Confirmation bias, the human tendency to ignore content that counters their prior beliefs, leads to individuals choosing to interact with like-minded individuals and ignoring interactions with the "opposite faction". This can compromise their open-mindedness, exacerbating and polarizing their views even more and ultimately leading to the formation and persistence of echo chambers.

Mathematical models of opinion formation, which incorporate psychological factors like cognitive biases, are often used to understand the effects of different kinds of biases on public opinions, especially political opinions. However, these models lack empirical validation. With the advent of the Internet and the rise of social media, human interactions are leaving an increasing digital footprint that can be exploited to study the dynamics of opinion formation and diffusion.

The purpose of this work is to empirically study the effects of political polarization on users' open-mindedness over time, particularly within the context of online social networks. The research focuses on a twenty-month discussion on r/politics, the largest political subreddit on Reddit, and aims to understand how open-mindedness is distributed within the population and over time and whether there is a significant difference in the level of open-mindedness in relation to political leaning. The study leverages the vast amount of big data traces left from online social networks to quantify socio-psychological and external drivers of opinion dynamics.

The results of the analysis suggest that different subpopulations' open-mindedness distributions are stable over time and statistically different, with Moderates/Neutrals and Republicans showing more nuanced close-mindedness patterns compared to Democrats. The research also highlights how Reddit users tend to be consistent in their open-mindedness attitude over time, showing, on average, low degrees of variance and dispersion.

### 8.1.1 Methodology: estimating open-mindedness on networks

In the following, we describe the first methodology developed to determine the level of open-mindedness - or bounded confidence - of users participating in an online discussion. The methodology is general and can be applied to different social networks and online contexts. The algorithm describing this methodology - as well as a detailed discussion of how it works - is provided below.

As already stated in the introduction to this chapter, in order to understand the levels of open-mindedness involved in the process of changing one's political leaning, we started by assuming a simple process of opinion evolution at the individual level, based on a very well known model of opinion formation [54]. In opinion dynamics models, agents update their opinions after interacting with their neighbors according to simple mathematical rules. For example, we recall that in [54], agents average their opinion with the opinion of their interacting peer, which is randomly chosen from the pool of their neighbors, if and only if their opinion distance is below a certain threshold representing the open-mindedness of the population (see Definition 11). The hypothesis that open-mindedness is a characteristic trait of an entire population and not a characteristic that varies from individual to individual is strong and probably unrealistic. For this reason, in the present work, we assume a Deffuant-like process of opinion update (i.e., users averaging their opinions with the opinions of their interacting partners in a pairwise fashion) and provide a data-driven time-aware estimate of individual-level open-mindedness. To estimate users' tendency to be influenced by their neighbors, we developed a simple approach (see Algorithm 3 for implementation details) that takes as input the weighted interaction network at time $t$ (a snapshot network, see Section 2.1.3 for more details) and the opinions of the agents at time $t$ and time $t+1$. In the estimation procedure, we select each node $u$ for which we have both opinions $x_u(t)$ and $x_u(t+1)$ (Algorithm 3 line 1) so that we can estimate how much the interactions that happened in that time step influenced the opinion change and therefore obtaining an estimate for the level

**Algorithm 3 Confidence bound estimation algorithm.**
$G_t$ = Weighted undirected interaction network at time t;
$V_t$ = set of nodes at time t;
$E_t$ = set of weighted edges at time t;
$x_u(t)$ = opinion of agent $u$ at time $t$;
$d_{u,v}$ = $|x_u(t) - x_v(t)|$ = opinion distance between $u, v \in V$ at time t;
$\widehat{CB}$ = estimated confidence bounds.

```
   if u ∈ V_{t+1} then
      Procedure to estimate x̂_u(t+1) and ĈB_u
3:    N_{u,t} = {v|(u,v) ∈ E_t}; n = |N_{u,t}|
      X_{N_{u,t}}[1...n] = Array of opinions of nodes v ∈ N_{u,t}
      if N_{u,t} ≠ ∅ then
6:       Sort X_{N_{u,t}}[1...n] by d_{u,v} in ascending order.
         X̂_u(t+1)[0...n] array of estimated opinions
         X̂_u(t+1)[0] = x_u(t)
9:       E = [0...n] array of estimation errors
         E[0] = 1.0
         for i=1; i=n; i++ do
12:         x_v = X_{N_{u,t}}[i]
            X̂_u(t+1)[i] = (X̂_u(t+1)[i−1] + x_v) / 2
            E[i] = |X̂_u(t+1)[i] − x_u(t+1)|
15:         min_e = E[n]
            for i=n; i=0; i– do
               e = E[i]
18:            if e ≤ min_e then
                  min_e = e
                  j = i
21:      x̂_u(t+1) = X̂_u(t+1)[j]
         ĈB = |x_u(t) − X_{N_{v,t}}[j]|
```

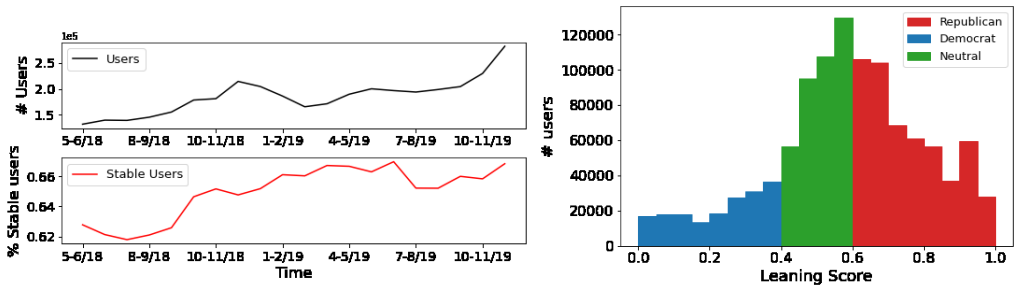
of bounded confidence. After selecting $u$, we order all the interacting partners (the neighbors of the node in the snapshot network) from the closer to the further by the opinion distance absolute value $d_{u,v}(t) = |x_u(t) - x_i(t)|$ (Algorithm 3 line 6). Then we compute an estimate $\widehat{x_u}(t+1)$ by iteratively averaging the new estimate with the interacting neighbors (Algorithm 3 line 13). The final estimated value $\widehat{x_u}(t+1)$ is the one that minimizes the error with respect to the real value $x_u(t+1)$ (Algorithm 3 lines 15-22). Finally, we compute the confidence bound as the distance in absolute value with the neighbor that represents the point of minimum in the estimation errors sequence (Algorithm 3 line 25). With the proposed approach, we can compute an estimate only for the subset of users present in two consecutive observations (i.e., months) and have at least one link (i.e., interaction) in the snapshot graph.

### 8.1.2 Data Collection.

For the purpose of this work, we built a dataset of political discussions on Reddit. Reddit is a popular social platform that allows users to post content to individual forums called subreddits, each dedicated to a specific topic. Such a categorized structure makes it easy to find users involved in specific debates. To assess the open-mindedness of users over time, we decided to select a quite controversial domain, i.e., politics. For a more detailed description of Reddit and its value for researchers, please refer to Section 4.2.

Among the thousands of subreddits talking about politics we choose `r/politics`[1], since it is the largest political subreddit and further, it is not aligned

[1]https://www.reddit.com/r/politics/ "`r/politics` is for news and discussion about US politics."




(A) (B)

FIGURE 8.1: a) Top: For each month, the number of users who participated in the debate. Bottom: For each month, the percentage of users that are stable across contiguous months. b) Authors' leaning distribution in the whole time period.

with a specific ideology but rather is visited regularly by users having different political beliefs. Notice that, as highlighted in the subreddit description, `r/politics` mainly refers to political discussion in the US. Thanks to the Pushshift API [20], we collected all posts and comments shared on the subreddit from May 2018 to December 2019, i.e., about one year and a half of Donald Trump's presidency - covering all discussions of 1,089795 users. As shown in Figure 8.1a, the number of users who participated in the debate tends to increase over time, and further, more than 60% of users are stable across contiguous months, meaning that they continue posting or commenting at least for two consecutive months. The code and the data used for this work are available in a dedicated GitHub repository[2].

### 8.1.3 Ideology Estimate.

To assess if the open-mindedness of users evolves over time, we have to establish the ideology of users in different time periods. Since we are dealing with users debating political issues in the US, we try to categorize them with respect to the US two-party system: *Republican* and *Democratic*. For such a purpose, we model the task of predicting users' political alignment as a text classification problem. In other words, we leverage users' posts to measure their degree of alignment with Republican and Democrat ideologies. To accomplish this task, we leverage an LSTM model that we trained on Reddit US political texts in previous works by Morini et al. [163, 164]. In detail, to train such a model, we defined a ground truth composed of Reddit posts belonging to subreddits known to be either Pro-Trump or Anti-Trump (i.e., `r/The_Donald` for the first group and `r/Fucktheatright` and `r/EnoughTrumpSpam` for the second). Accordingly, we modeled the text classification task as a binary problem. During model selection, we perform a 3-fold Cross-Validation trying different hyper-parameters configurations and obtaining the best performances on the validation set (i.e., the average accuracy of 82.9%) using GloVe word embeddings and 128 LSTM units [3].

Consequently, we apply the model to the `r/politics` dataset in order to infer posts leaning for all the population. Notice that we apply it separately for each month in order to assess users' ideology stability over time. For each textual content, we obtain model predictions ranging from 0 to 1 (i.e., the model confidence), where 1

[2] https://github.com/ValentinaPansanella/OpenMindednessReddit.git

[3] For further details on the model selection and evaluation steps, the reader should refer to our previous works [163, 164]

| $N$ | $N_R$ | $N_D$ | $N_N$ | $E$ | $\langle k \rangle$ | $r$ |
|---|---|---|---|---|---|---|
| 183296 | 88285 | 36385 | 58626 | 1271080 | 6.97 | 0.020 |

TABLE 8.1: Network statistics averaged across the 20 considered months: number of users $N$, divided in Republican $N_R$, Democrat $N_D$ and Neutral $N_N$, number of edges $E$, network average degree $\langle k \rangle$, and network assortativity $r$ with respect to the political leaning.

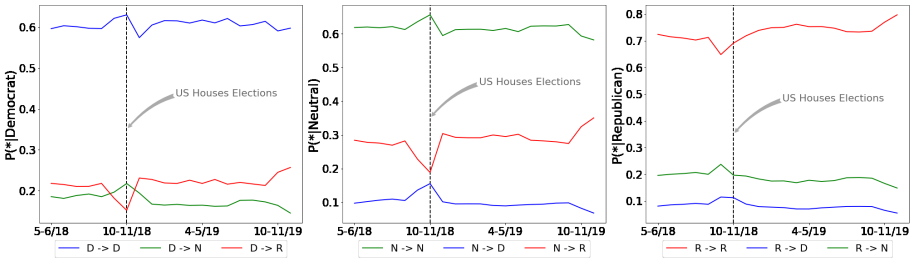



FIGURE 8.2: For each ideology, users' transition probabilities over contiguous months.

means that the post aligns itself with Republican (specifically Pro-Trump) ideologies while 0 with Democrats (specifically Anti-Trump) ones. Then, for each user $u$ who participated in the debate in each month $m$, we compute his *leaning score*, $L_{u,m}$, as the average value of his monthly posts leaning as follows:

$$L_{u,m} = \frac{\sum_{p \in P_{u,m}} \text{Prediction\_Score}(p)}{|P_{u,m}|} \tag{8.1}$$

where $P_{u,m}$ is the set of posts shared by a user $u$ in each month $m$. Since we are interested in assessing if users with polarized opinions tend to move to more moderate positions, we discretized such leanings into three intervals: *Democrat* if $L_{u,m} \leq 0.4$; *Republican* if $L_{u,m} \geq 0.6$; while *Neutral* if $0.4 < L_{u,m} < 0.6$. By adding a third label, we make sure to capture users with highly polarized ideologies. Figure 8.1b shows the authors' leaning score distribution for the whole time period obtained by averaging users' monthly scores. Such a distribution confirms what was observed in a recent work [52] that focuses the analysis on the `r/politics` subreddit too: Republican users (530,909) outnumber Democrat ones (185,256) and Neutrals users (373,630) show a tendency towards republicans beliefs.

### 8.1.4 Network Definition.

Given the users for which we inferred their ideology, we define their interaction networks for each month to take into account the evolution of leanings in time. The resulting networks have Reddit users as node sets, $V$, and as edges the set of pair $(u, v) \in V$ for which a reply of $u$ to a $v$'s post or a comment exists. We set each edge weight to represent the total number of comments exchanged between two users. Also, we label users (i.e., nodes) with their discretized *leaning score* $L_{u,m}$ (i.e., Republican, Neutral, and Democrat). In Table 8.1, we provide the main averaged network statistics across the 20 considered months.

### 8.1.5 Ideology Stability over Time.

As a preliminary analysis, we try to understand how users' ideologies evolve. In detail, we are interested in exploring if users are stable and consistent with the same

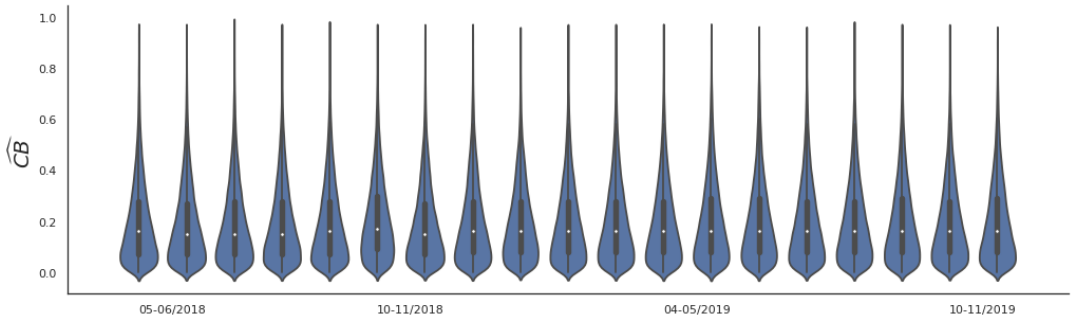



FIGURE 8.3: Estimated open-mindedness $\widehat{CB}$ distribution from May-2018 to December-2019. Distributions of the estimated confidence bound $\widehat{CB}$ over the whole time period. All distributions are positively skewed and constant over time.

ideology or instead tend to change opinions according to specific events. We model such an issue in terms of transition probabilities for such a purpose. In other words, for each user, we compute his probability $p_{ij}$ to move from *state* $i$ to *state* $j$ over contiguous months. Notice that, in this scenario, *state* stands for the user's ideology (i.e., Republican, Neutral or Democrat).

In Figure 8.2, we show, for each ideology, the probability of users to change or remain in their states (i.e., leanings) over contiguous months. At first glance, we can observe similar behavior for all the political ideologies: users tend to be rooted in their position over months, with a probability to remain in their state ranging around 0.60 for Democrat and Neutral leanings and around 0.70 for Republican. Moreover, Republican users - when changing their state - foster a more neutral position instead of moving to Democrat beliefs (differently from Democrats). These attitudes reflect the extreme polarization that characterizes the US political debate in the Trump Era [17]. However, for all the considered ideologies, we also notice an evident fluctuation of opinions in favor of Democrats and Neutrals around November 2018, precisely when (i.e., 6 November) the Democratic Party won control of the US House from the Republican Party. Indeed, this was a decisive moment in which Democrats won the seats needed to take the House after capturing districts where President Trump was unpopular.

### 8.1.6 open-mindedness distributions

Figure 8.3 underlines that the distribution of the estimated open-mindedness $\widehat{CB}$ is stable during the considered time period in the analyzed discussion. Moreover, it also highlights that the majority of the analyzed users is "close-minded", i.e., their confidence bound is $\widehat{CB} \leq 0.2$, which is considered to be a sufficient condition for the population to become polarized in the long term, according to [54]. This means that most of the users participating in this discussion can hardly be influenced by neighbors holding distant opinions, even if they interact with these users during the considered time period, like in the case of this discussion, where the network has a low assortativity with respect to the political leaning (see network assortativity in Table 8.1). However, we can also see that the distributions at each time step have a very high variance, allowing the presence of individuals having a level of open-mindedness close to 1.0 - indicating that some of the users can also be influenced by people holding very different opinions and changing their expressed political leaning accordingly. The distribution of the estimated confidence bound for the opinions is less skewed between October and November 2018, i.e., around the US House of

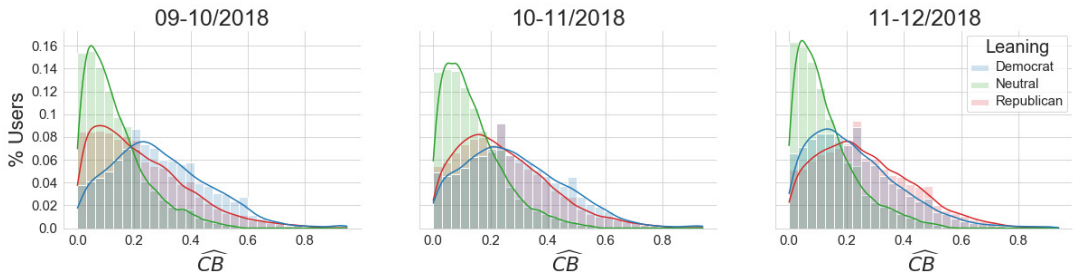
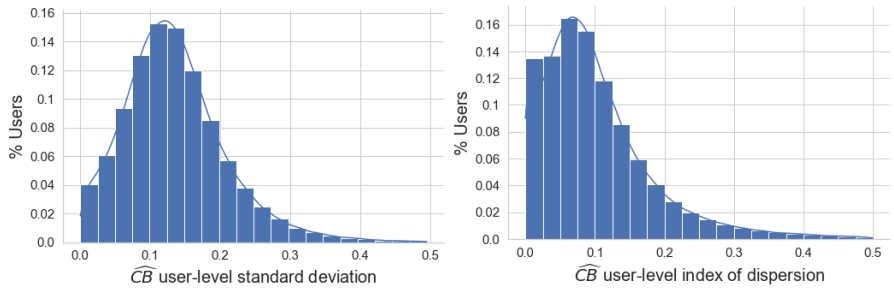



FIGURE 8.4: Estimated open-mindedness distribution in the period September - December 2018. Distributions of the estimated confidence bound $\widehat{CB}$ for the different political leanings: Democrat (blue), Neutral (green), and Republican (red) from September to December 2018.

Representatives elections. Such behavior confirms the data-driven analysis based on transition probabilities performed in Section 8.1.5: in this time window, Republicans, which normally have a highly skewed distribution, seemed to be somehow more open-minded, and their average confidence bound is higher.

In Figure 8.4 we reported the estimated distribution for Democrat, Neutral, and Republican users (we took the orientation at time $t$ and estimated the confidence bound $\widehat{CB}$ between time $t$ and $t+1$). Only three months are present in this figure, i.e., the three months around the US House election, but conclusions still hold for the other months considered in this work. We can see from Figure 8.4 that there are differences in the distribution of open-mindedness when we consider political orientation. Both neutral and republican distributions are positively skewed, and the distribution has a long right tail. Also, the Democrats' distribution has a right tail, but much less than the others. From our analysis, it appears that Neutral individuals are also the most close-minded, while Democrats have a wider range of confidence bound levels. Their distribution is not as skewed as the others, and many users have a very high level of bounded confidence, i.e., they change their opinion significantly over contiguous time periods under the influence of their neighbors. Republicans, like Neutrals, have a positively skewed confidence bound distribution, even with a higher average confidence bound. We performed a 2-sample KS-test comparing the distributions of the estimated confidence bound for each political leaning (e.g., Democrat vs. Republican $\widehat{CB}$) within each time step obtaining a p-value $\approx 0.0$, supporting the conclusions that distributions are different for the three political leanings. Finally, while we can say that population-level open-mindedness is reasonably constant over time, i.e., we have the same mix of open-minded and close-minded individuals participating in the discussion, we do not have information about how variable open-mindedness may be at the individual level. In this analysis, each user may have a different value of open-mindedness at each time step, making not only the overall distribution heterogeneous but also the distribution at the individual level. To understand how much individual bounded confidence may vary, we computed the standard deviation of our estimate for each observed user - Figure 8.5a. Reddit users' open-mindedness tendencies appear stable in time, showing a characteristic low standard deviation, $\sigma$. Such a result is also confirmed by the distribution of the Fano dispersion index (Figure 8.5b) - i.e., the ratio between variance $\sigma^2$ and mean value, $\mu$, of the estimated individual open-mindedness scores. The observed Fano values, prevalently distributed in the range $0 < \frac{\sigma^2}{\mu} < 1$, identify under-dispersed behaviors, thus expressing consistent patterns of stability.




(A) (B)

FIGURE 8.5: Users' open-mindedness stability analysis. (A) Distribution of individuals' open-mindedness standard deviations; (B) Distribution of individuals' open-mindedness dispersion indexes (variance over mean value).

### 8.1.7 Conclusions

When applied to study the political debate that took place on Reddit during the first two years of the Trump presidency, our proxy unveiled the existence of characteristic distributions for well-defined sub-populations: Moderates/Neutrals and Republicans being more close-minded on average than Democrats. The proposed longitudinal analysis also unveiled that the observed Reddit users tend to maintain a stable behavior for what concerns their open-mindedness, exhibiting low variance and underdispersion.

Indeed the current study, like all empirical ones, suffers from limitations. In particular, it leverages a data-driven estimate of the political leaning that can be subject to errors and cannot be fully validated on ground truth external data. Moreover, although Reddit users tend to be particularly involved in political discussions, the population variability in time and the sparsity of observation data do not allow us to estimate the open-mindedness of less active individuals. For each month, we were not able to estimate the open-mindedness of $40 - 50\%$ of nodes since we have no information on the opinion at time $t + 1$ or the user does not have neighbors on the network at time $t$.

As future works, we plan to enhance the proposed estimation procedure, allowing for asymmetric open-mindedness. Additionally, we will investigate the interplay of open-mindedness and known polarization phenomena (e.g., the presence of echo chambers) in order better to characterize the role of different individuals in their emergence.

## 8.2 Estimating Open-Mindedness in Controversial Reddit Discussions: A Comparative Network and Hypergraph Approach

In Section 8.1, we developed a data-driven time-aware methodology that estimates users' open-mindedness, starting from users' interactions represented as networks, to overcome the lack of data-driven approaches to calibrate and validate opinion dynamics models. However, in many online contexts (e.g., Reddit), people mainly participate in group discussions, which could be better captured by exploiting higher-order structures, e.g., hypergraphs. As more thoroughly explained in Section 2.1.2,

these structures model user interactions as subsets of nodes called hyperedges instead of pairs.

In this Section, we want to expand the methodology described in Section 8.1.1 and understand the differences emerging from using the two different frameworks to model online discussions. In particular, we want to estimate user-level open-mindedness using the algorithm in Section 8.2.1 and compare the differences in the estimation of opinions and open-mindedness with respect to the other case. Moreover, we want to compare the different discussions to account for topic-specific behaviors, which could not emerge from the work in Section 8.1 due to the employment of a single dataset.

In order to do this, we chose three different datasets, which are all built on "controversies", in which we expect users to have similar behaviors from a cognitive perspective in their process of interactions and opinion change.

### 8.2.1 Methodology: estimating open-mindedness on hypergraphs

The current lack of data-driven approaches that validate models on real data has led us to develop a data-driven time-aware methodology [178] that estimates users' open-mindedness, starting from users' interactions represented as networks, described in Section 8.1.1. In many online contexts, such as Reddit, Twitter, Facebook Groups, and other platforms, people primarily engage in group discussions, which can be better understood by examining higher-order structures rather than pairwise interactions.

One crucial characteristic of higher-order structures is peer pressure, a potent social force that can mould the behavior and beliefs of the group's constituents. Although an individual may feel free to express their genuine thoughts during a one-on-one interaction, the collective viewpoint can significantly impact a group setting, causing individuals to conform even if it means suppressing their own views. This kind of "spiral of silence" phenomenon can amplify certain opinions and suppress others, leading to polarization and echo chambers.

In a pairwise framework, this phenomenon would be less observable as it is a collective behavior rather than an individual one.

In order to account for such characteristics, we extended the approach described in Section 8.1.1 to hypergraphs.

The primary prerequisite for initiating the estimation process remains unaltered: a node must be present in two consecutive time intervals, as stated in line 1 of Algorithm 4. In the context of networks, the opinions of nodes adjacent to $u$ during the time interval $t$ were gathered in the array $N_{u,t}$, as specified in line 3 of Algorithm 3. To adapt Algorithm 3 for hypergraphs, certain modifications are necessary. In particular, it is essential to collect the political inclinations of all nodes encompassed within the hyperedges of the star ego network of node $u$. To achieve this, an array of arrays, $X_{C_{u,t}}$, is defined in line 5 of Algorithm 4, with each nested array corresponding to a hyperedge in the star ego-network of $u$. These arrays contain the political leanings of all neighboring nodes within the respective hyperedge. The cardinality of the array $X_{C_{u,t}}$ is equivalent to the number of hyperedges in which node $u$ is present. To derive the array $\overline{X}_{C_{u,t}}$, the average political leaning of all nodes included in each hyperedge containing node $u$ at time $t$ is computed, as outlined in line 8 of Algorithm 4. This array is analogous to the $X_{N_{u,t}}$ array employed in line 3 of Algorithm 3. The array will be utilized to estimate the opinion of the node in the subsequent interval $t+1$. For each value in the array $\overline{X}_{C_{u,t}}$, the opinion of node $u$ at time $t+1$ is iteratively estimated (as shown in line 16 of Algorithm 4) by averaging the current $\widehat{x_u}(t+1)$

with each element of $\overline{X}_{C_{u,t}}$. The optimal estimated opinion, $\widehat{x_u}(t+1)$, is determined by minimizing the discrepancy between the estimated value $\widehat{x_u}(t+1)$ and the observed opinion at the following interval $x_u(t+1)$, as demonstrated in lines 20 to 22 of Algorithm 4. The confidence bound, $\widehat{CB}$, is computed as the absolute value of the difference between node $u$'s opinion and the mean value of the hyperedge opinions that yielded the smallest estimation error, as indicated in line 28 of Algorithm 4.

**Algorithm 4 Estimating open-mindedness on hypergraphs.**
$V_t$ = set of nodes at time $t$;
$H_t$ = set of hyperedges at time $t$;
$C_{u,t}$ = set of hyperedges of the node $u$ star ego-network;
$x_u(t)$ = opinion of agent $u$ at time $t$;
$\widehat{CB}$ = estimated confidence bounds.

```
   if u ∈ V_{t+1} then
2:    Procedure to estimate x̂_u(t+1) and ĈB_u
      |C_{u,t}| = n;
4:    if C_{u,t} ≠ ∅ then
         X_{C_{u,t}}[1...n] Array of opinions of the node u star ego-network
6:       X̄_{C_{u,t}}[1...n] Array for the average opinion for each hyperedge
         for i=0; i=n; i++ do
8:          X̄_{C_{u,t}}[i] = (Σ C_{u,t}[i]) / |C_{u,t}[i]|
         X̂_u(t+1)[0...n] Array of estimated opinions
10:      X̂_u(t+1) = x_u(t)
         E = [0...n] Array of estimation errors
12:      E[i] = 1.0
         for i=1; i=n; i++ do
14:         x_v = X̄_{C_{u,t}}[i]
            X̂_u(t+1)[i] = (X̂_u(t+1)[i−1] + x_v) / 2
16:         E[i] = |X̂_u(t+1)[i] − x_u(t+1)|
            min_e = E[n]
18:         for i=n; i=0; i- - do
               e = E[i]
20:            if e ≤ min_e then
                  min_e = e
22:               j = i
         x̂_u(t+1) = X̂_u(t+1)[j]
24:      ĈB = |x_u(t) − X_{N_{v,t}}[j]|
```

The primary distinction between the two methods lies in the manner in which the estimated opinion of node $u$ is calculated. In Algorithm 3, the estimated opinion is derived from the mean of the individual neighbor's opinion that resulted in the smallest error. Conversely, in Algorithm 4, the estimation is based on the average opinion of the interaction context of node $u$ that produced the lowest error. In the former approach, interactions occur pairwise between neighboring nodes, whereas in the latter, interactions transpire simultaneously among nodes connected by the same hyperedge. Consequently, the hyperedge is regarded as the interaction context in the second method. This distinction highlights the unique characteristics of each algorithm and their respective approaches to estimating node opinions within different network structures.

### 8.2.2 Datasets

Given that the emergence of online social network platforms has revolutionized political expression and participation, enabling individuals to access news and information, and engage in active discussions with other users, we decided to continue

using Reddit, a popular platform for opinion sharing and political debates, as the primary source of data for this study (see Section 4.2 for more details).

The data used in this research was previously retrieved for a study on network echo chambers by Morini et al. [163] using the Pushshift API [20].

The collected posts cover the period from January 2017 to July 2019, which corresponds to the first two and a half years of Donald Trump's presidency. This period was characterized by highly polarized political discussions, reflecting the divisive nature of the Trump Era. Additionally, the dataset includes the mid-term elections held on November 6, 2018, where the Democratic party gained control of the U.S. House of Representatives while the Republican party maintained control of the U.S. Senate.

To gain insights into different sociopolitical issues and study how interactions vary based on the discussion topic, the data was collected from several subreddits via Reddit List and divided into three distinct datasets:

- **Gun Control**: This dataset focuses on discussions related to gun control policy. It includes posts and comments from subreddits both in favor of and against gun legalization.
- **Minorities Discrimination**: This dataset encompasses discussions on discrimination against minorities. It includes users with conservative ideas as well as those advocating for gender, sexual, and racial equality.
- **Politics**: This dataset comprises posts and comments from general political discussions, providing a broader view of the U.S. sociopolitical landscape.

**Opinion estimation.** To label each user with a political opinion based on the text of their posts and comments, a BERT text classifier [56] was trained. A ground truth dataset was created by selecting subreddits known to have highly polarized positions, such as "r/The_Donald," "r/FuckTheRight," and "r/EnoughTrumpSpam." The BERT model was then applied to the Gun Control, Minorities Discrimination, and Politics datasets, generating a prediction score ranging from 0 to 1. A score of 1 represents a pro-Trump ideology, 0 represents an anti-Trump ideology, and scores in between indicate a neutral or moderate stance. For the purpose of this work, we assumed that a pro-Trump ideology aligns with Republican political leaning, an anti-Trump ideology aligns with Democrat political leaning, and a neutral stance represents a moderate position. Differently from Section 8.1, data aggregation was conducted by semester instead of monthly. This was done because there are more users present in contiguous semesters than in contiguous months, which leads to more robust estimations of user opinions due to increased user activity, such as multiple posts. Moreover, this allowed us to estimate the open-mindedness of a higher percentage of users, since we need two contiguous opinions for the methodology detailed in Section 8.2.1. The users' six-monthly political leaning scores, denoted as $L_{u,s}$, are computed as the average value of the monthly post leaning for each user $u$ participating in the discussions during each semester $s$. The calculation is performed using Equation (8.1), aggregating over a semester $s$ instead of a month $m$. Here, $P_{u,s}$ represents the set of posts shared by user $u$ in each semester $s$. The resulting $L_{u,s}$ values range from 0 to 1 and are discretized into three intervals: $L_{u,s} \leq 0.4$ for Democrats, $L_{u,s} \geq 0.6$ for Republicans, and $0.4 \leq L_{u,s} \leq 0.6$ for Moderates, as already done in [164]. This third label is used to identify users with a political position that does not align exclusively with either the Democratic or Republican party. After the data collection and pre-processing, interaction structures

were built from the Reddit discussions, using both snapshot graphs and snapshot hypergraphs modeling frameworks and a semester as a time window to aggregate data.

### Graph definition

Each snapshot network denoted as $G_s$, is constructed as an undirected weighted graph. The graph consists of a set of nodes, $V_s$, where each node represents a user who participated in the discussion at time $s$. Users participate in the discussion by posting or commenting on other users' posts. An edge $(u, v)$ is present between two users if they have commented on each other's posts. Additionally, each node in the network has an attribute representing their political leaning, denoted as $L_{u,s}$.

The weight of each edge $(u, v)$ is represented by an integer number, $w_s(u, v)$, indicating the number of interactions (reciprocal comments) between users $u$ and $v$ during semester $s$.

| **Gun control dataset** | | | | | | |
|---|---|---|---|---|---|---|
| Interval | Nodes | Edges | Avg Cl. Coeff. | Avg Deg. | N. Comp. | Density |
| 01-07 2017 | 833 | 4044 | 0.1898 | 9.7095 | 8 | 0.01167 |
| 07-12 2017 | 847 | 3925 | 0.1687 | 9.268 | 6 | 0.010955 |
| 01-07 2018 | 1054 | 3942 | 0.1363 | 7.4800 | 4 | 0.0071 |
| 07-12 2018 | 985 | 3478 | 0.1090 | 7.0619 | 16 | 0.0072 |
| 01-07 2019 | 1046 | 3601 | 0.0904 | 6.8853 | 6 | 0.006589 |

TABLE 8.2: Snapshots graph properties for the Gun Control dataset for each time window

| **Minority dataset** | | | | | | |
|---|---|---|---|---|---|---|
| Interval | Nodes | Edges | Avg Cl. Coeff. | Avg Deg. | N. Comp. | Density |
| 01-07 2017 | 1040 | 3765 | 0.214 | 7.240 | 3 | 0.00697 |
| 07-12 2017 | 1004 | 3465 | 0.200 | 6.902 | 5 | 0.00688 |
| 01-07 2018 | 1170 | 3832 | 0.185 | 6.550 | 5 | 0.0056 |
| 07-12 2018 | 1113 | 3594 | 0.161 | 6.458 | 6 | 0.0058 |
| 01-07 2019 | 1126 | 3405 | 0.154 | 6.048 | 5 | 0.00538 |

TABLE 8.3: Snapshots graph properties for the Minority dataset for each time window

| **Politics dataset** | | | | | | |
|---|---|---|---|---|---|---|
| Interval | Nodes | Edges | Avg Cl. Coeff. | Avg Deg. | N. Comp. | Density |
| 01-07 2017 | 917 | 2525 | 0.165 | 5.507 | 4 | 0.0060 |
| 07-12 2017 | 746 | 1816 | 0.149 | 4.869 | 10 | 0.0065 |
| 01-07 2018 | 825 | 2179 | 0.138 | 5.282 | 5 | 0.0064 |
| 07-12 2018 | 775 | 1787 | 0.124 | 4.612 | 5 | 0.0059 |
| 01-07 2019 | 686 | 1411 | 0.098 | 4.114 | 11 | 0.0060 |

TABLE 8.4: Snapshots graph properties for the Politics dataset for each time window

In tables 8.2, 8.3, and 8.4, we summarized the networks' characteristics for each dataset during the five intervals.

During this period, the number of users increases, allowing insights into the evolution of users' political leanings $L_{u,s}$. More than 60% of users are present across consecutive months, which is essential for computing the open-mindedness level.

For the Gun Control and Minorities datasets, the number of nodes steadily increases over time, indicating increased user engagement in discussions related to these controversial topics during the Trump era. Significant events, such as the Parkland mass shooting in February 2018, may have influenced user participation, as evidenced by a sudden increase in users in the third semester of the Gun Control dataset.

For what concerns the opinion distribution of the population within each semester, the Gun Control and Minorities datasets demonstrate a roughly uniform distribution, with a good proportion of Republicans, Democrats, and Moderates. In the Politics dataset, instead, the distribution is skewed, exhibiting a low proportion of Republican users participating in the discussions, compared to the other two political leanings.

#### Hypergraph Definition

The same datasets introduced in the previous paragraph were also modeled using the (snapshot) hypergraph framework, as presented in Section 2.1.2. The nodes and their opinion variables remain the same, as do the temporal intervals of six months each, resulting in five hypergraph snapshots.

Leveraging the original temporal networks, the hypergraph structure is inferred by means of all the maximal cliques [66]. In this way, the hyperedge becomes the context of interaction of the nodes included in it, highlighting the multiplicity of points of view that simultaneously participate in the discussion. Most hyperedges include few nodes, indicating that the contexts of interactions analyzed are mostly small. The majority of nodes are included in less than 25 hyperedges for the Gun Control dataset and less than 10 for the other two datasets.

Not every hyperedge links the same number of nodes, so an insight into the distribution of the hyperedge dimension is shown in figure 8.6 respectively for Gun Control, Minorities, and Politics dataset.

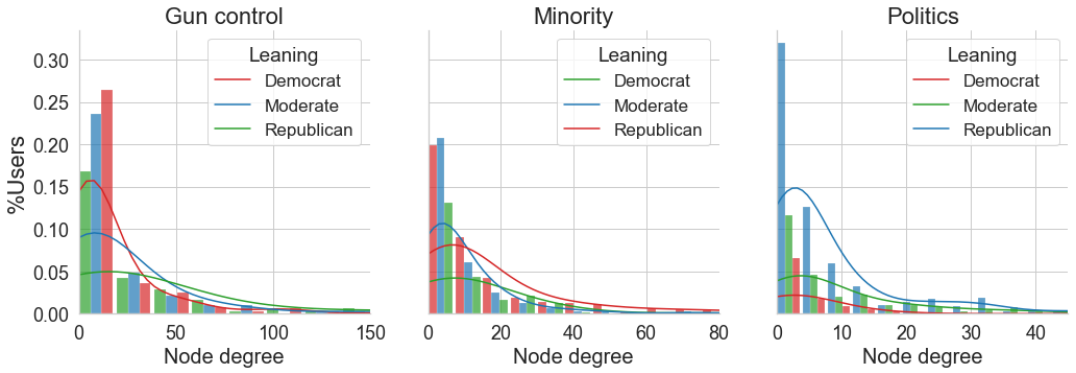



FIGURE 8.6: Node degree distributions for each dataset (Gun control, Minority and Politics) in the hypergraphs.

### 8.2.3 Results

In the following, we present and discuss the outcomes gained by employing the methodology described in Section 8.1.1 and Section 8.2.1 on each of the datasets presented in Section 8.2.2.

From our analysis, it emerges that regardless of the type of underlying structure employed to model discussions and regardless of the topic discussed (at least

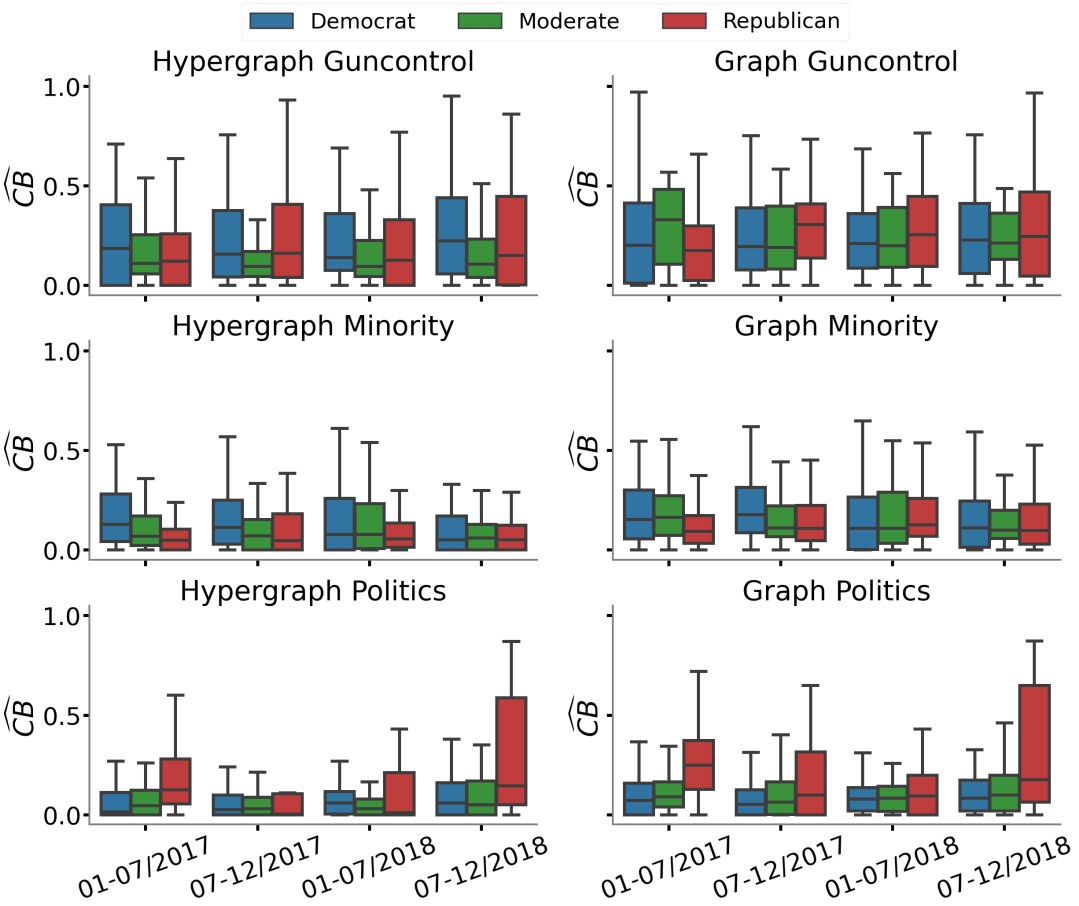



FIGURE 8.7: **open-mindedness Distributions**. In the left column, we plot results obtained with the methodology described in Section 8.2.1 on the data described in Section 8.2.2. In the right column, we plot results obtained with the methodology described in Section 8.1.1 on the data described in Section 8.2.2. Colors refer to political leaning as estimated with the procedure of Section 8.1.3: blue indicates Democrats, green Moderates, and red Republicans. Each row represents a dataset (from top to bottom): Gun Control, Minority Regulation, and Politics.

while remaining in the realm of controversial discussions), Reddit users are overall "close-minded" – which may drive the users towards more polarized point of views – although this does not emerge from the temporal evolution of the opinion distribution in these datasets.

As we can see from Figure 8.7, the median $\widehat{CB}$ is steadily below 0.35, which is the known threshold for *consensus*, for what concerns the model as described in [54].

More specifically, in the Gun Control dataset, Moderates are the most close-minded group in both configurations, but their $\widehat{CB}$ is lower in the hypergraph, with a mean value of 0.2 and a narrower interquartile range. This tendency of having a lower - median - $\widehat{CB}$ in the hypergraph configuration is present in all datasets. The lowest (median) $\widehat{CB}$ values in the graph framework are always from Democrats or Republicans, while in the hypergraph, the lowest values are from the Moderates. In the Minority dataset, only Republicans show similar behaviors in both configurations. Democrats are the least close-minded, followed by Republicans and then Moderates. The hypergraph configuration shows a higher $\widehat{CB}$ level. In the Politics dataset, the trends of the three political leanings are very similar in both configurations, i.e., the median $\widehat{CB}$ presents a u-shaped evolution in the four semesters, and in

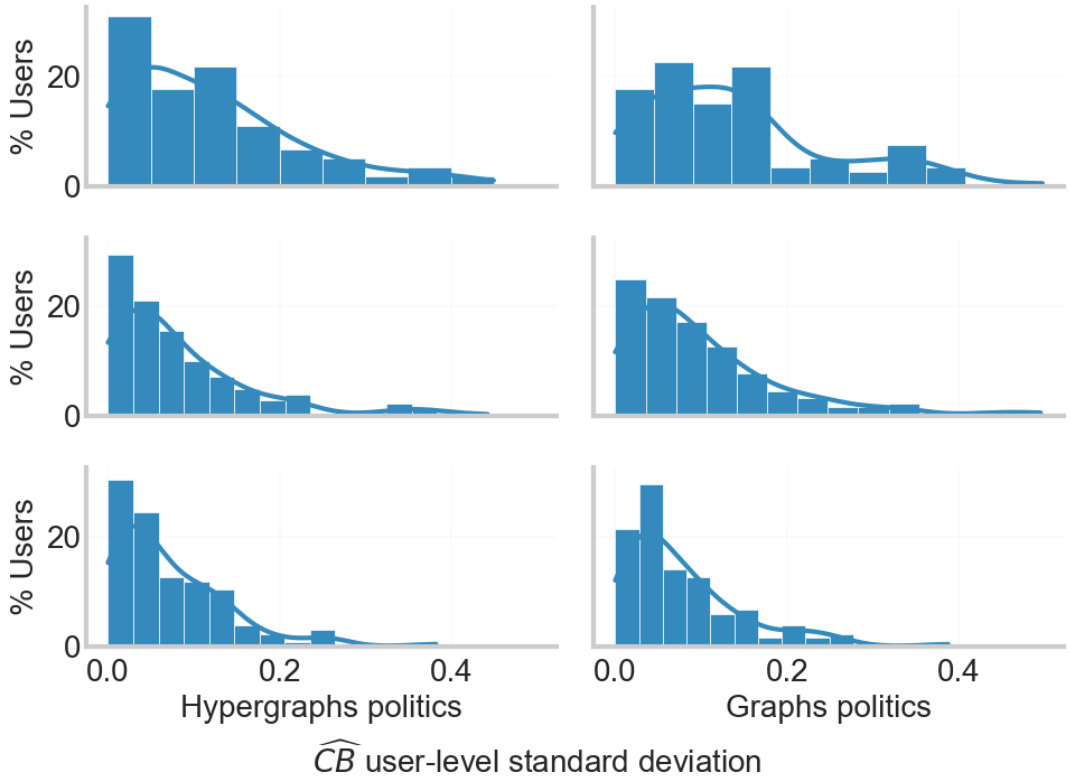



FIGURE 8.8: User-level standard deviation of $\widehat{CB}$ for both Hypergraph (left column) and Graph (right column) frameworks.

particular, in the last semester, the third quartile is above 0.5. This may be due to the fact that – in this specific subreddit – the republican population is a minority and, over time, is less and less strongly anchored to its political leaning and "opens its mind" to evolve towards more moderate positions. The hypergraph configuration produces higher $\widehat{CB}$ values and a higher level of estimated opinion error. Overall, the hypergraph configuration tends to produce higher values of $\widehat{CB}$ and estimation error compared to the network configuration. This is attributed to the context of group interactions modeled by the hyperedges, which amplify the influence of nodes with more distant opinions during the opinion estimation process.

The higher-order approach confirms results obtained in Section 8.1 on the stability of the distributions over time: despite changing the modeling framework and the time-window for aggregating data, we can see from Figure 8.7 (left column) that subpopulations (by political leaning) tend to show the same open-mindedness distribution over time.

One may ask if this stable behavior depends on user-level stability or not: as showed in Figure 8.8 users have a low temporal variability with respect to $\widehat{CB}$, which suggests that users with a higher threshold tend to maintain such level of open-mindedness over time, and the same applies to users with a lower threshold. This confirms the insights described in Section 8.1.

An interesting finding of this analysis is that open-mindedness and homophily – computed as the mean variance of the opinions on the hyperedges each user participates in – are not correlated (the Pearson correlation coefficient is 0.04), i.e., engaging with a more diverse context does not affect having a higher (or lower) open-mindedness.

This encourages the idea that open-mindedness – or, one may say, influenciability – is a cognitive feature that plays a role in the process of opinion change of individuals, regardless of their tendencies to select narrower or broader contexts of interactions, i.e., having homophilic or heterophilic discussions, in a given period of time.

A necessary step in the procedures in Sections 8.1.1 and 8.2.1 is the estimation of the user's opinion at time $t+1$ to compute $\widehat{CB}$. In the analysis across various datasets, we observed a distinct pattern of estimation errors between the hypergraph and graph frameworks. The estimation error of $x_u(t+1)$ is computed as the absolute value of the difference between the node $u$ observed opinion and the estimated node opinion (Algorithm 3 line). The hypergraph framework, which inherently captures higher-order interactions, yielded higher estimation errors across all datasets: guncontrol (0.084), minority (0.065), and politics (0.046). On the other hand, the graph framework, which encapsulates pairwise interactions, demonstrated lower estimation errors for the same datasets: guncontrol (0.061), minority (0.054), and politics (0.04). This discrepancy in estimation errors could be attributed to the inherent complexity of the hypergraph framework, which may not always provide a more accurate representation of the unfolding of social influence in online discussions. The graph framework, despite its higher simplicity, appears to be more effective in this context, possibly due to its ability to capture the essential dynamics of the discussions without the added complexity of higher-order interactions. However, it is important to note that these results do not undermine the potential of the hypergraph framework in modeling complex social phenomena. Rather, they highlight the need for further research to refine the application of hypergraphs in the field of computational social sciences, particularly in the context of online social networks, opinion dynamics, and polarization.

### 8.2.4 Conclusions

In this Section, we used a data-driven approach to estimate the level of individual open-mindedness in different controversial debates on Reddit about U.S. politics. The data collected covers three main discussion topics: gun control legislation, minority discrimination, and general socio-politics arguments. The data is organized in two different network structures: graphs (pairwise interactions) and hypergraphs (high-order interactions). The work aims to estimate the confidence bound $\widehat{CB}$, which represents the maximum opinion distance between users' opinions for active interaction between them, from real interaction data. The results show that the majority of Reddit users are closed-minded, with the Moderates being the most closed-minded subpopulation and Republicans surprisingly showing an increase in open-mindedness in the case of discussing Political topics. Most results confirm insights already obtained in Section 8.1.

## 8.3 From models to data to models: understanding real opinion dynamics on Twitter

This last section focuses on a key aspect that brings together the culmination of our research efforts. The aim here was to bridge a noticeable gap in the existing research by creating a synthesis of the entire doctoral study.

Our main goal was to gain a comprehensive understanding of how opinions evolve in complex social networks, with a focus on the interplay between algorithmic biases, bounded confidence, and real network structures.

In Chapter 5, we developed a novel opinion dynamics model, namely the *ABMM Model*. We also came up with a novel approach to estimate an important parameter of this model – namely the confidence bound ($\epsilon$) – with a particular emphasis on making it adaptable to individual users in Section 8.1.1.

Despite employing the proposed methodology for descriptive purposes – i.e., to analyze the user-level distribution of open-mindedness in online political discussions – the potential impact of such an approach was not fully exploited in the previously described case studies.

In this last section, we decided to use the methodology in Section 8.1.1 to calibrate a heterogeneous version of the *ABMM Model*.

Our case study revolved around the Euro 2020 "taking the knee" controversy – a polarizing subject widely discussed in mass media and social networks, with distinct narratives. Our investigation aimed to unravel the phenomenon where intensely polarized debates gradually transition into depolarization over time.

### 8.3.1 Dataset and Methods

The dataset used in this study spans approximately one month, from June 10th to July 13th, during which the EURO2020 matches were played. To focus our analysis on relevant conversations, we applied hashtag-based filtering, targeting discussions related to Italy's matches, the tournament itself, and the topic of taking the knee. This filtering process yielded a collection of 38,908 tweets authored by 16,235 unique users.

We adopted a hashtag-based approach to infer Twitter users' opinions regarding taking the knee during EURO 2020. A manual annotation process was employed to classify 2,304 hashtags from the dataset. Each hashtag was assigned a numerical value based on its alignment with the pro or against stance, with $\pm 3$ indicating a clear position, $\pm 1$ indicating a close association, and 0 assigned to neutral or irrelevant hashtags. We calculated the non-neutral hashtag values within each tweet by averaging its classification value ($C_t$). Similarly, for each user ($u$), we computed their overall classification value ($C_u$) by averaging the classification values of their tweets. To facilitate integration with our opinion dynamics model, the initial pro/against scores, ranging from $-3$ to 3, were normalized to a range of $[0, 1]$. Additionally, we discretized the leanings into three categories: "Pro" (if $C_u \leq 0.4$), "Against" (if $C_u \geq 0.6$), and "Neutral" otherwise, encompassing users with highly polarized viewpoints.

From the collected data, we constructed an undirected attributed temporal network, where nodes represent users and edges capture their interactions, including retweets, mentions, quotes, and replies. The resulting network comprises 15,378 nodes and 36,496 edges. To serve as initial and final states for validating our model, we divided the network into two snapshots: the first corresponding to the group stage and round-of-16, and the second representing the period from the quarterfinals to the final. This division was chosen based on specific reasons that will be further specified. As our model does not consider the temporal evolution of links, we retained only the nodes present in both snapshots. The temporal element was disregarded, resulting in two undirected snapshot networks: $G_0$, with nodes labeled according to their leaning in the first period, and $G_1$, with nodes labeled according to their leaning in the second period. This simplification aligns with our model's

assumption of a static network. The two snapshot graphs consist of 2,925 users (approximately 20% of the total) and 9,081 edges. Notably, the giant connected component comprises 2,894 nodes and 9,054 edges.

**Experiments on real data**

The experiments were carried out with the following parameters:

- The underlying network structure is $G$: each node $u$ in the interaction network is an agent $i$ and each leaning $C_u$ in $G_0$ is an opinion $x_i$ with $x_i \in [0, 1]$.
- We tested both homogeneous and heterogeneous bounded confidence levels. For homogeneous values we considered$\epsilon \in \{0.2, 0.3, 0.4\}$; for heterogeneous values, each agent $i$ is assigned with a level of bounded confidence $\epsilon_i$ obtained applying the procedure in Section 8.1.1 Algorithm 3 to $G_0, G_1$.
- The parameter $p_m$ takes values of either 0.0 (in the absence of mass media, the model becomes the Algorithmic Bias Model with heterogeneous $\epsilon$) or 0.5.
- The parameter $\gamma$ varies in the range of [0.0, 1.5] with increments of 0.5; for $\gamma = 0.0$, we obtain the Deffuant-Weisbuch model with heterogeneous $\epsilon$.
- The parameter $\mu$ is set to 0.5, i.e., when two agents interact, they adopt their average opinion.
- The maximum number of iterations is set at $10^5$.
- Simulations terminate when the maximum opinion change remains below a threshold of 0.01 for at least 500 consecutive iterations.

We performed a comprehensive analysis to examine the influence of different scenarios on opinion evolution. Our investigation encompassed five distinct media landscapes:

- One mass media with opinion $x_m = avg(pro) = 0.28$
- One mass media with opinion $x_m = avg(neutral) = 0.49$
- One mass media with opinion $x_m = avg(against) = 0.87$
- Two mass media with opinions $x_{m1} = avg(pro) = 0.28$ and $x_{m2} = avg(against) = 0.87$
- Three mass media with opinions $x_{m1} = avg(pro) = 0.28$, $x_{m2} = avg(against) = 0.87$ and $x_{m3} = avg(neutral) = 0.49$

Since in these experiments every agent $i$ has a different level of bounded confidence $\epsilon_i$, to account for parameter heterogeneity, we applied the opinion change rule of the Algorithmic Bias Model with Mass Media in the following way:

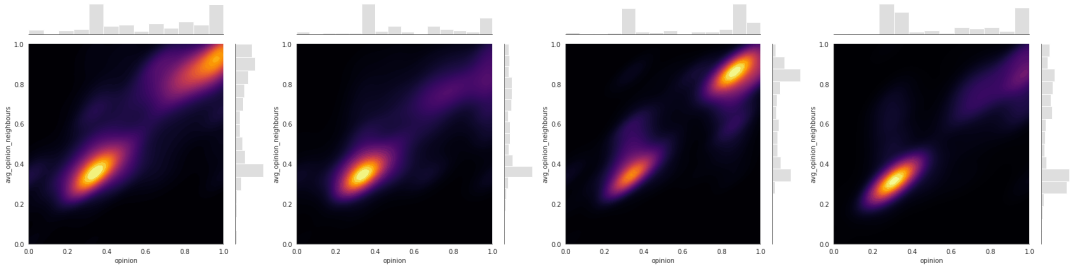
- if $d_{ij} < \epsilon_i$ $x_i(t+1) = (x_i(t) + x_j(t))/2$
- if $d_{ij} < \epsilon_j$ $x_j(t+1) = (x_i(t) + x_j(t))/2$
- if if $d_{ij} > \epsilon_i$ & $d_{ij} > \epsilon_j$ nothing happens

i.e., a heterogeneous version of the baseline model.

Since we performed only one run per scenario, it is not feasible to compute the same metrics used in the mean-field analysis. Therefore, we choose to compare the simulation outcomes under various conditions on the real network with the actual opinion values in $G_1$. This allows for a direct assessment of the simulation results against the empirical opinion data at the specified time point. Specifically, we conducted one simulation for each scenario and compared the results with $G_1$ by examining the final states. To assess polarization and the presence of echo chambers in both real data and simulation outcomes, we adopted the approach presented in [43].

These plots provide insights into the formation of echo chambers within an interaction network by analyzing the behavior of individual nodes in relation to their neighbors' behavior. As in [43], we measure polarization in our simulation results based on the correlation between a user's leaning and the average leaning of their nearest neighbors (ego network).

### 8.3.2 Results: Algorithmic bias depolarizes discussion on EURO2020 "taking the knee" controversy



(A) $G_0$ (B) $G_1$ (C) $p_m = 0.5$, $\gamma = 1.5$, $x_m = 0.87$ (D) $p_m = 0.5$, $\gamma = 1.5$, $x_m = 0.28$

FIGURE 8.9: **Joint distribution of the opinion of users and average leaning of their neighborhood.** We display the first snapshot $G_0$ (initial matches) (A); the second snapshot $G_1$ (quarter-finals to final) (B); the final state of the simulation of the Algorithmic Bias Model with Mass Media and Heterogeneous Confidence Bounds with $p_m = 0.5$, $\gamma = 1.5$ and $x_m = 0.87$(C); and the final state of the simulation of the Algorithmic Bias Model with Mass Media and Heterogeneous Confidence Bounds with $p_m = 0.5$, $\gamma = 1.5$ and $x_m = 0.28$(D).

Despite trying to capture possible real dynamics with mathematical models of opinion formation, such synthetic settings may fail to capture peculiar characteristics of real networks, e.g., scale-free degree distributions and modular structures, but also polarized initial conditions, which may characterize discussions around controversial topics. Such diverse conditions may lead to different conclusions than the ones obtained in the mean-field case. For this reason, we exploited an empirical network collected from Twitter during EURO2020, where Italian users expressed their stances on the controversy around taking the knee in favor of the Black Lives Matter protests [31]. We simulated our model using this network as a starting condition (both topology and initial opinion distribution) for different values of the model's parameters.

Our findings suggest that consensus may be reached in the final state when considering a homogeneous confidence threshold in scenarios with no media present or only a single media source. Even if such results are not averaged over multiple runs, these results may imply that scale-free degree distributions and modular topologies

enhance consensus when the population has a homogeneous level of bounded confidence that is not lower than 0.2. However, an exception arises when there are no media sources, and a parameter value of $\gamma$=1.5 is applied. In this case, the final opinion distribution becomes fragmented, characterized by two main clusters centered around the average leaning of the "pro" faction and the average leaning of the "against" faction. In this case, the bias may be too strong for users to converge toward a common opinion. When two polarized media sources are introduced, opinions are concentrated around a moderate opinion in the final distributions. It exhibits a Gaussian shape, suggesting that the population tends to converge towards a common opinion in this case too. However, the presence of polarized media may keep users leaning toward more extreme positions. Adding a "moderate" media to this scenario, our observations reveal that the final opinion distribution remains symmetric and peaked around the center of the opinion spectrum. However, the distribution variance decreases compared to the previous scenario, i.e., people tend to homologate even more around a single opinion value, and variability is reduced. Furthermore, as the bias ($\gamma$) increases, the variance continues to decrease, and for $\gamma = 1.0$, a single main opinion cluster emerges in the final state. Nevertheless, if the bias increases, e.g., $\gamma = 1.5$, the final distribution splits into distinct opinion clusters centered around the media opinion. Moreover, since assumptions of homogeneous parameters are considered unrealistic, we exploited a methodology developed in Section 8.1.1 to estimate user-level open-mindedness ($\epsilon_i$) and simulated a heterogeneous extension of our model.

As displayed in Figure 8.9(A), users were embedded into echo chambers around pro and against stances on the discussion during the first two matches. However, when considering the period from the quarter-finals to the final (Figure 8.9(B)), the same users are mainly clustered around positions in favor of kneeling, and polarization appears to be reduced. Simulations of our model, which exploits the first network as initial conditions of the simulations and accounts for heterogeneous levels of the confidence threshold estimated from the data according to the procedure in Section 8.1.1, appear to confirm some of the insights offered by the mean-field analysis on the complete network with homogeneous parameters. The main conclusion that also holds in a real setting is that the algorithmic bias favors opinion fragmentation but, in doing so, helps to reduce the level of polarization of the network (see Figure 8.9(C) and (D)) when there is an external source (or even a highly influential user) promoting one stance over the other. However, the setting that better captures the real opinion evolution can be seen in Figure 8.9(D), where a stubborn agent is promoting a fixed opinion aligned with the stance in favor of players "taking the knee". However, in Figure 8.9(C), where the media is aligned with the opposite stance, the community that becomes less polarized is the other one, differently from the real situation.

### 8.3.3 Conclusions

Real network structures, characterized by scale-free degree distributions, modular structures, and polarized initial conditions, clearly impact the results of the dynamics of the present model. When open-mindedness is homogeneous across the population, users tend to converge towards a single opinion value, which depends on the initial average opinion and the opinion promoted by a single media. When the media landscape is more heterogeneous, i.e., media supporting two opposite stances, the population still tends to conform to a moderate stance. However, the final distribution has a higher variability, with some users maintaining more extreme leanings.

Such variability is reduced when the media landscape actively promotes more moderate stances. In the case study, cognitive biases do not play a role in the result of the dynamics, while the role of the algorithmic bias remains the same as in the baseline model. However, when inferring open-mindedness levels from empirical data and using the real distribution of the parameter to simulate the model, results show a final polarization distribution closer to the real ones, and the depolarizing role of the algorithmic bias emerges. Specifically, the real final state is well approximated by the setting where there is a recommender system biasing interactions and a mass media promoting an opinion aligned with the "pro-taking-the-knee" faction.

# Chapter 9

# Conclusion

In the conclusion of this thesis, we circle back to our original aim: to unravel the complex interplay between biases (Section 4.2) and network effects (Chapter 2) in the process of opinion formation and diffusion (Chapter 3) within online social networks. Our approach to reaching this goal was twofold.

In all Chapters of Part II, we developed new **models** of opinion dynamics, tailored for the specific characteristics of such online environments and their possible polluted nature. We simulated the long-term outcomes of these dynamics. In particular, we focused on studying the interplay of algorithmic bias, skewing interactions towards like-minded individuals, with complex network structures, to understand if scale-freeness or mesoscale structures impact the process of opinion formation differently (Chapter 6). The novelty of this work is not on the developed model – which is the same as [204] – but the analysis of the role of the underlying structure and the study of the role of initial condition (of both opinions and structure) on Polarization and Echo Chamber phenomena.

In Chapter 7, we encompassed the idea of an inevitable interplay between the process unfolding on the population (opinion evolution) and the dynamics *of* the population (the evolution of the social network structure), especially online social networks, allowing greater freedom to users with respect to the offline world, are characterized by bubbles of similar individuals connecting and interacting. The *AAB Model* simulates this mechanism with a simple rewiring rule: when two agents interact and "disagree", they can rewire their link toward a like-minded individual. The same Chapter also explores the role of peer-pressure in such biased environments. The *ABSC Model* incorporates the idea of a majority pressuring a single individual to conform using the Simplicial Complex framework as the underlying structure, where network triads become 2-simplexes. Unsurprisingly, this higher pressure towards conformity enhances consensus despite the polarizing power of the biases interplaying in the process. Although purely theoretical, such models incorporate elements of realism that can help researchers, policymakers, and platform owners reason about the unintended harmful effects caused by online behaviors and adopt countermeasures to avoid them.

In Chapter 5, we developed an extension of [204], namely the *ABMM Model*, to investigate the role of external agents, possibly (in)voluntarily steering the dynamics towards a desired goal. We analyzed different media landscape and their effects on the final state of the population (employing a mean-field approach in the present Chapter); the results led us to re-discuss the positive/negative meanings normally given to the role of bias (cognitive and algorithmic) and consensus/polarization/fragmentation states, showing how by changing external conditions, a previously undesirable outcome can become desirable. Despite the novelty of this work, a necessary advancement is to investigate the impact of different types

of media sources on opinion formation and dissemination. This thesis has demonstrated the strong influence of media on individual opinions. However, not all media sources are created equal. Future research could explore how the credibility, bias, or type of media source (e.g., mainstream media vs. social media) affects the opinion dynamics in online environments.

The role of social AI in shaping online opinion dynamics is a critical finding of this thesis. This underscores the need for future research on developing AI algorithms that promote diversity of opinions and prevent the formation of echo chambers. This could involve exploring AI bias mitigation techniques, such as fairness-aware machine learning or differential privacy. Additionally, there is a need for methods to evaluate the fairness and transparency of AI algorithms. How can we ensure that these algorithms are not inadvertently contributing to the polarization of opinions?

Part III surveys methodologies developed using an existing opinion dynamics model (the Deffuant-Weisbuch model [54]) to estimate user-level time-aware distributions of open-mindedness in real online discussions. In Chapter 8, we detail two different methodologies for the Graph (Section 8.1.1) and Hypergraph (Section 8.2.1) frameworks, and we show three different case studies on different datasets from Reddit and Twitter to understand the generalizability of our results and provide more robust conclusions.

The first two case studies respond to a desire to analyze the role of open-mindedness within polarized discussions on notoriously controversial issues. Specifically, we decided to study discussions of a political nature on Reddit (focusing on the US) and to go on to understand the distribution of open-mindedness within such populations, modeling the interactions (post-comments) as both networks and hypergraphs. The results obtained are, of course, not generalizable – they depend on the platform, topic, time window, and even the methodology chosen – however, they lay a first building block for the development of methods that take advantage of the simplicity and interpretability of opinion dynamics models to interpret real phenomena. The last section of this part (Section 8.3) closes the loop that forms the idea behind this thesis: being able to create a feedback-loop between models and data to try to explain real phenomena. Again, clearly, we cannot say that the results are generalizable beyond the specific case study, but – building on the methodology developed in Section 8.1 and the model developed in Chapter 5 – we simulated a heterogeneous version of the *ABMM Model* using as initial conditions those of a real Twitter discussion around the protests concerning the Black Lives Matter movement during the 2020 Europeans. Simulations of the model calibrated with real data show that the presence of a recommendation algorithm and an external force proposing a single opinion are the most likely conditions explaining the real evolution of the dynamics.

While this thesis work presents innovative and strong elements that contribute to advancing opinion dynamics models and studying human behavior in online social environments, it is important to acknowledge that there are also some limitations. In addition to the specific limitations discussed in each chapter, we will briefly outline some general limitations that apply to the entire work. It is worth noting that while data-driven elements are present, this work primarily focuses on modeling a specific class of models. This decision was made because these models are particularly suited to the goals of the work. However, it is important to acknowledge that there is no empirical evidence to suggest that they are the best possible models for studying offline and online human behavior. A comparison with other classes of models may offer a more comprehensive perspective, allowing the reader to distinguish which

results are generalizable and which are dependent on the specific characteristics of the model. One potential direction for future research is to delve deeper into the role of individual characteristics in shaping online opinion dynamics. While this thesis has focused on the collective behavior of individuals in online environments, there is a need to understand how individual differences contribute to these dynamics. For instance, how do personality traits, cognitive biases, or demographic factors influence an individual's susceptibility to conform to the dominant opinion? How do these individual characteristics interact with the social and technological factors identified in this thesis? It is important to acknowledge that the thesis presents data-driven work. However, the framework for calibrating and validating the models through real data presented may benefit from further development to increase its depth and generalizability to other models and case studies. This aspect needs fundamental improvement. One limitation of data-driven studies is that the evolution of a population's views is a slow process (beyond special cases such as might have been the one presented in Section 8.3), whose effects may only be observed in the long run. How do opinions change as individuals interact with different people or are exposed to different information over time? How do echo chambers and polarization develop and persist in the long run? Conducting longitudinal studies on the evolution of opinions rather than limiting ourselves to relatively short periods may be necessary. Longitudinal studies can be essential to understanding the long-term effects of changes in opinion. However, collecting and analyzing such a large amount of data can pose inherent difficulties.

In conclusion, this thesis has comprehensively explored opinion dynamics in online environments. It has revealed the complex interplay of individual behaviors, social interactions, and AI algorithms in shaping these dynamics. Understanding these processes will be crucial in fostering healthy and productive online discourse as we navigate the digital age. The journey may be challenging, but the potential rewards - a more inclusive, diverse, and balanced online world - are worth the effort.

# Bibliography


[1] Mohammad Afshar and Masoud Asadpour. "Opinion Formation by Informed Agents". In: *Journal of Artificial Societies and Social Simulation* 13.4 (2010), p. 5. DOI: 10.18564/jasss.1665.

[2] Wasim Ahmed, Peter A Bath, and Gianluca Demartini. "Using Twitter as a data source: An overview of ethical, legal, and methodological challenges". In: *The ethics of online research* 2 (2017), pp. 79–107.

[3] Alya Alaali, Maryam a. Purvis, and Bastin Tony Roy Savarimuthu. "Vector Opinion Dynamics: An Extended Model for Consensus in Social Networks". In: *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* 3 (2008), pp. 394–397. DOI: 10.1109/WIIAT.2008.377.

[4] Rex N Ali and Saswati Sarkar. "Impact of opinion dynamics on the public health damage inflicted by COVID-19 in the presence of societal heterogeneities". In: *Frontiers in Digital Health* 5 (2023), p. 1146178.

[5] Frédéric Amblard and Guillaume Deffuant. "The role of network topology on extremism propagation with the relative agreement opinion dynamics". In: *Physica A: Statistical Mechanics and its Applications* 343 (2004), pp. 725–738. DOI: 10.1016/j.physa.2004.06.102.

[6] Sinan Aral, Lev Muchnik, and Arun Sundararajan. "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks". In: *Proceedings of the National Academy of Sciences* 106.51 (2009), pp. 21544–21549.

[7] Hannah Arendt. "Freedom and politics". In: *Freedom and serfdom: An anthology of Western thought* (1961), pp. 191–217.

[8] Solomon E Asch. "Studies of independence and conformity: I. A minority of one against a unanimous majority." In: *Psychological monographs: General and applied* 70.9 (1956), p. 1.

[9] Robert Axelrod. "The Dissemination of Culture: A Model with Local Convergence and Global Polarization". In: *The Journal of Conflict Resolution* 41.2 (1997), pp. 203–226. URL: http://www.jstor.org/stable/174371.

[10] Chris Bail. *Breaking the social media prism: How to make our platforms less polarizing*. Princeton University Press, 2022.

[11] Christopher A. Bail et al. "Exposure to opposing views on social media can increase political polarization". In: *Proceedings of the National Academy of Sciences* 115.37 (2018), pp. 9216–9221. DOI: 10.1073/pnas.1804840115.

[12] E. Bakshy, S. Messing, and L. A. Adamic. "Exposure to ideologically diverse news and opinion on facebook". In: *Science* 348.6239 (2015), pp. 1130–1132.

[13] A. Barabasi. "The origin of bursts and heavy tails in human dynamics". In: *Nature* 435 (2005), pp. 207–211.

[14] Albert-Lazlo Barabási. *Network Science*. Cambridge University Press, 2016.

[15] Albert-Lazlo Barabási and Réka Albert. "Emergence of scaling in random networks". In: *Science* 286.5439 (1999), pp. 509–512.

[16] Hugo Barbosa et al. "Human mobility: Models and applications". In: *Physics Reports* 734 (2018), pp. 1–74.

[17] L. M. Bartels. "Partisanship in the trump era". In: *The Journal of Politics* 80.4 (2018), pp. 1483–1494.

[18] Federico Battiston et al. "Networks beyond pairwise interactions: structure and dynamics". In: *Physics Reports* 874 (2020), pp. 1–92.

[19] Michael Batty. "The size, scale, and shape of cities". In: *Science* 319.5864 (2008), pp. 769–771.

[20] Jason Baumgartner et al. "The Pushshift Reddit Dataset". In: *Proceedings of the International AAAI Conference on Web and Social Media* (2020), pp. 830–839.

[21] Alessandro Bellina et al. "Effect of collaborative-filtering-based recommendation algorithms on opinion polarization". In: *Physical Review E* 108.5 (Nov. 2023). ISSN: 2470-0053. DOI: 10.1103/physreve.108.054304.

[22] FP Bianchi and S Tafuri. "A public health perspective on the responsibility of mass media for the outcome of the anti-COVID-19 vaccination campaign: the AstraZeneca case." In: *Annali di Igiene, Medicina Preventiva e di Comunita* 34.6 (2022).

[23] Stefano Boccaletti et al. "The structure and dynamics of multilayer networks". In: *Physics reports* 544.1 (2014), pp. 1–122.

[24] A. Bornhäuser, J. McCarthy, and S. Glantz. "German tobacco industry's successful efforts to maintain scientific and political respectability to prevent regulation of secondhand smoke". In: *Tobacco Control* 15 (2006), e1–e1.

[25] Laurent Boudin and Francesco Salvarani. "Opinion dynamics: kinetic modelling with mass media, application to the Scottish independence referendum". In: *Physica A: Statistical Mechanics and its Applications* 444 (2016), pp. 448–457.

[26] Engin Bozdag. "Bias in algorithmic filtering and personalization". In: *Ethics and information technology* 15 (2013), pp. 209–227.

[27] David Broockman and Joshua Kalla. "The manifold effects of partisan media on viewers' beliefs and attitudes: A field experiment with Fox News viewers". In: *OSF Preprints* 1 (2022), pp. 1–42.

[28] Heather Z Brooks and Mason A Porter. "A model for the influence of media on the ideology of content in online social networks". In: *Physical Review Research* 2.2 (2020), p. 023041.

[29] Jane D Brown and Kim Walsh-Childers. "Effects of media on personal and public health". In: *Media effects* (2002), pp. 463–498.

[30] Axel Bruns. "Filter bubble". In: *Internet Policy Review* 8.4 (2019).

[31] Chiara Buongiovanni et al. "Will You Take the Knee? Italian Twitter Echo Chambers' Genesis During EURO 2020". In: *Complex Networks and Their Applications XI: Proceedings of The Eleventh International Conference on Complex Networks and Their Applications: COMPLEX NETWORKS 2022–Volume 1* (2023), pp. 29–40.

[32] Christian Burgers and Anneke De Graaf. "Language intensity as a sensationalistic news feature: The influence of style on sensationalism perceptions and effects". In: *Communications-The European Journal of Communication Research* 38.2 (2013), pp. 167–188.

[33] T. Carletti et al. "How to make an efficient propaganda". In: *EPL* 74 (2006), pp. 222–228.

[34] Michael X Delli Carpini and Scott Keeter. *What Americans know about politics and why it matters*. Yale University Press, 1996.

[35] C. Castellano, M. A. Muñoz, and R. Pastor-Satorras. "Nonlinear q-voter model." In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 80 4 Pt 1.4 (2009), p. 041129.

[36] Pew Reseach Center. "Political polarization in the american public". In: *Ann Rev Polit Sci* (2014).

[37] David Chavalarias, Paul Bouchaud, and Maziyar Panahi. "Can Few Lines of Code Change Society? Beyond fack-checking and moderation: how recommender systems toxifies social networking sites". In: *arXiv preprint arXiv:2303.15035* (2023).

[38] Ge Chen et al. "Convergence properties of the heterogeneous Deffuant–Weisbuch model". In: *Automatica* 114 (2020), p. 108825.

[39] Cynthia Chew and Gunther Eysenbach. "Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak". In: *PloS one* 5.11 (2010), e14118.

[40] Sandeep Chowdhary et al. "Simplicial contagion in temporal higher-order networks". In: *Journal of Physics: Complexity* 2.3 (2021), p. 035019.

[41] Nicolas H Christianson, Ann Sizemore Blevins, and Danielle S Bassett. "Architecture and evolution of semantic networks in mathematics texts". In: *Proceedings of the Royal Society A* 476.2239 (2020), p. 20190741.

[42] Weiqi Chu and Mason A Porter. "Non-Markovian models of opinion dynamics on temporal networks". In: *SIAM Journal on Applied Dynamical Systems* 22.3 (2023), pp. 2624–2647.

[43] Matteo Cinelli et al. "The echo chamber effect on social media". In: *Proceedings of the National Academy of Sciences* 118.9 (2021), e2023301118. DOI: 10.1073/pnas.2023301118. URL: https://www.pnas.org/content/118/9/e2023301118.

[44] Federico Cinus et al. "The effect of people recommenders on echo chambers and polarization". In: *Proceedings of the International AAAI Conference on Web and Social Media* 16 (2022), pp. 90–101.

[45] Salvatore Citraro. "Feature-rich Networks: When Topology meets Semantics". PhD thesis. University of Pisa, Department of Computer Science, 2023.

[46] Peter Clifford and Aidan Sudbury. "A Model for Spatial Conflict". In: *Biometrika* 60.3 (1973), pp. 581–588. DOI: 10.2307/2335008. URL: http://www.jstor.org/stable/2335008.

[47] Auguste Comte. *The positive philosophy of Auguste Comte*. C. Blanchard, 1855.

[48] M. Conover et al. "Political polarization on twitter". In: *Proceedings of the International AAAI Conference on Web and Social Media* (2011), pp. 89–96.

[49] Rosaria Conte et al. "Manifesto of computational social science". In: *The European Physical Journal Special Topics* 214 (2012), pp. 325–346.

[50] Michele Coscia. "The atlas for the aspiring network scientist". In: *arXiv preprint arXiv:2101.00863* (2021).

[51] Felipe Cucker, Steve Smale, and Ding-Xuan Zhou. "Modeling language evolution". In: *Foundations of Computational Mathematics* 4 (2004), pp. 315–343.

[52] Abir De et al. "Learning and forecasting opinion dynamics in social networks". In: *Advances in neural information processing systems* (2016), pp. 397–405.

[53] Guillaume Deffuant, Sylvie Huet, and Frédéric Amblard. "An individual-based model of innovation diffusion mixing social value and individual benefit". In: *American journal of sociology* 110.4 (2005), pp. 1041–1069.

[54] Guillaume Deffuant et al. "Mixing beliefs among interacting agents". In: *Advances in Complex Systems* 3.01n04 (2000), pp. 87–98.

[55] M. Degroot. "Reaching a Consensus". In: *Journal of the American Statistical Association* 69 (1974), pp. 118–121.

[56] Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[57] Andrea Di Benedetto et al. "Media preference increases polarization in an agent-based election model". In: *Physica A: Statistical Mechanics and its Applications* 626 (2023), p. 129014.

[58] Júlia Domingo, Guillaume Diss, and Ben Lehner. "Pairwise and higher-order genetic interactions during the evolution of a tRNA". In: *Nature* 558.7708 (2018), pp. 117–121.

[59] Yucheng Dong et al. "A survey on the fusion process in opinion dynamics". In: *Inf. Fusion* 43 (2018), pp. 57–65.

[60] Tim Donkers and Jürgen Ziegler. "De-Sounding Echo Chambers: Simulation-Based Analysis of Polarization Dynamics in Social Networks". In: *Available at SSRN 4437898* (2023).

[61] James N Druckman and Michael Parkin. "The impact of media bias: How editorial slant affects voters". In: *The Journal of Politics* 67.4 (2005), pp. 1030–1049.

[62] Elizabeth Dubois and Grant Blank. "The echo chamber is overstated: the moderating effect of political interest and diverse media". In: *Information, communication & society* 21.5 (2018), pp. 729–745.

[63] Erick Elejalde, Leo Ferres, and Rossano Schifanella. "Understanding news outlets' audience-targeting patterns". In: *EPJ Data Science* 8.1 (2019), pp. 1–20.

[64] Paul Erdős and Alfred Rényi. "On random graphs. I". In: *Publicationes Mathematicae* 6 (1959), pp. 290–297.

[65] Giorgio Fagiolo, Marco Valente, and Nicolaas J Vriend. "Segregation in networks". In: *Journal of economic behavior & organization* 64.3-4 (2007), pp. 316–336.

[66] Andrea Failla, Salvatore Citraro, and Giulio Rossetti. "Attributed Stream Hypergraphs: temporal modeling of node-attributed high-order interactions". In: *Applied Network Science* 8.1 (2023), pp. 1–19.

[67] Juan Fernández-Gracia, Vıctor M Eguıluz, and Maxi San Miguel. "Update rules and interevent time distributions: Slow ordering versus no ordering in the voter model". In: *Physical Review E* 84.1 (2011), p. 015103.

[68] Myra Marx Ferree et al. "Four models of the public sphere in modern democracies". In: *Theory and society* 31.3 (2002), pp. 289–324.

[69] Giulia Ferro, Valentina Pansanella, and Giulio Rossetti. "Higher-Order Estimate of Open Mindedness in Online Political Discussions". Master's Thesis. University of Pisa, 2023.

[70] Leon Festinger. "A theory of cognitive dissonance". In: 2 (1957).

[71] Peter Fischer et al. "The theory of cognitive dissonance: State of the science and directions for future research". In: *Clashes of Knowledge* (2008), pp. 189–198.

[72] Jason D Flatt, Yll Agimi, and Steve M Albert. "Homophily and health behavior in social networks of older adults". In: *Family & community health* 35.4 (2012), pp. 312–321.

[73] Santo Fortunato. "Universality of the threshold for complete consensus for the opinion dynamics of Deffuant et al." In: *International Journal of Modern Physics C* 15 (2004), pp. 1301–1307.

[74] M. Franke and R. Rooij. "Strategies of Persuasion, Manipulation and Propaganda: Psychological and Social Aspects". In: *Models of Strategic Reasoning* (2015).

[75] Jan-Philipp Fränken and Toby Pilditch. "Cascades across networks are sufficient for the formation of echo chambers: An agent-based model". In: *Journal of Artificial Societies and Social Simulation* 24.3 (2021).

[76] M. Franklin, Michael Marsh, and Lauren Mclaren. "Uncorking the Bottle: Popular Opposition to European Unification in the Wake of Maastricht". In: *Journal of Common Market Studies* 32 (1994), pp. 455–472.

[77] N. E. Friedkin. "A formal theory of social power". In: *Journal of mathematical sociology* 12 (1986), pp. 103–126.

[78] Noah Friedkin and Eugene Johnsen. "Social Influence Networks and Opinion Change". In: *Advances in Group Processes* 16 (1999).

[79] Noah E Friedkin. *A structural theory of social influence.* Cambridge University Press, 1998.

[80] Friedman R. Friedman M. *Tyranny of the status quo*. San Diego,Harcourt Brace Jovanovich, 1984.

[81] Karen Friend and David T Levy. "Reductions in smoking prevalence and cigarette consumption associated with mass-media campaigns". In: *Health education research* 17.1 (2002), pp. 85–98.

[82] Wyndol Furman and Valerie A Simon. "Homophily in adolescent romantic relationships". In: *Understanding peer influence in children and adolescents* (2008), pp. 203–224.

[83] Serge Galam. "Minority opinion spreading in random geometry." In: *Eur.Phys. J. B* 25.4 (2002), pp. 403–406.

[84] Franco Galante et al. "Modeling communication asymmetry and content personalization in online social networks". In: *Online Social Networks and Media* 37 (2023), p. 100269.

[85] Yérali Gandica et al. "Continuous opinion model in small-world directed networks". In: *Physica A-statistical Mechanics and Its Applications* 389 (2010), pp. 5864–5870.

[86] Floriana Gargiulo and Yerali Gandica. "The role of homophily in the emergence of opinion controversies". In: *arXiv preprint arXiv:1612.05483* (2016).

[87] Floriana Gargiulo, Stefano Lottini, and Alberto Mazzoni. "The saturation threshold of public opinion: are aggressive media campaigns always effective?" In: *arXiv preprint arXiv:0807.3937* (2008).

[88] Kiran Garimella et al. "Reducing controversy by connecting opposing views". In: *Proceedings of the tenth ACM international conference on web search and data mining* (2017), pp. 81–90.

[89] V. Garimella et al. "Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship". In: *Proceedings of the 2018 World Wide Web Conference* (2018).

[90] Anna Gausen, Wayne Luk, and Ce Guo. "Using agent-based modelling to evaluate the impact of algorithmic curation on social media". In: *ACM Journal of Data and Information Quality* 15.1 (2022), pp. 1–24.

[91] Y. Ge et al. "Understanding echo chambers in e-commerce recommender systems". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (2020), pp. 2261–2270.

[92] Santiago Torres Gil and Damián H. Zanette. "Coevolution of agents and networks: Opinion spreading and community disconnection". In: *Physics Letters A* 356 (2006), pp. 89–94. URL: https://api.semanticscholar.org/CorpusID:18046128.

[93] Michelle Girvan and Mark EJ Newman. "Community structure in social and biological networks". In: *Proceedings of the national academy of sciences* 99.12 (2002), pp. 7821–7826.

[94] C. A. Glass and D. H. Glass. "Opinion dynamics of social learning with a conflicting source". In: *Physica A-statistical Mechanics and Its Applications* 563 (2021), p. 125480.

[95] Sergiy Gnatyuk et al. "Quantitative evaluation method for mass media manipulative influence on public opinion". In: *CEUR Workshop Proceedings* (2019).

[96] Mark Granovetter. "Threshold models of collective behavior". In: *American journal of sociology* 83.6 (1978), pp. 1420–1443.

[97] Mark S Granovetter. "The strength of weak ties". In: *American journal of sociology* 78.6 (1973), pp. 1360–1380.

[98] Thilo Gross, Carlos J Dommar D'Lima, and Bernd Blasius. "Epidemic dynamics on an adaptive network". In: *Physical review letters* 96.20 (2006), p. 208701.

[99] Beniamino Guerra et al. "Dynamical organization towards consensus in the Axelrod model on complex networks". In: *Physical Review E* 81.5 (2010), p. 056105.

[100] L. Guo and X. Cai. "Continuous opinion dynamics in complex networks". In: *Communications in Computational Physics* 5.5 (2009), pp. 1045–1053.

[101] Felix Hamborg, Karsten Donnay, and Bela Gipp. "Automated identification of media bias in news articles: an interdisciplinary literature review". In: *International Journal on Digital Libraries* 20.4 (2019), pp. 391–415.

[102] William Hart et al. "Feeling validated versus being correct: a meta-analysis of selective exposure to information." In: *Psychological bulletin* 135.4 (2009), p. 555.

[103] Daniel B. M. Haun and Michael Tomasello. "Conformity to peer pressure in preschool children." In: *Child development* 82 6 (2011), pp. 1759–67.

[104] R. Hegselmann and U. Krause. "Opinion dynamics and bounded confidence: models, analysis and simulation". In: *J. Artif. Soc. Soc. Simul.* 5 (2002).

[105] Abigail Hickok et al. "A Bounded-Confidence Model of Opinion Dynamics on Hypergraphs". In: *arXiv preprint arXiv:2102.06825* 21.1 (2021), pp. 1–32.

[106] Thomas T. Hills. "The Dark Side of Information Proliferation". In: *Perspectives on Psychological Science* 14 (2019), pp. 323–330.

[107] Thomas Hobbes. *Leviathan; or, The matter, forme, & power of a common-wealth, ecclesiasticall and civill*. London, Printed for A. Crooke, 1651. URL: http://archive.org/details/leviathan00hobba.

[108] Eileen A. Hogan. "The Attention Economy: Understanding the New Currency of Business". In: *Academy of Management Perspectives* 15 (2001), pp. 145–147.

[109] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. "Stochastic blockmodels: First steps". In: *Social networks* 5.2 (1983), pp. 109–137.

[110] R. A. Holley and T. M. Liggett. "Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model". In: *The Annals of Probability* 3.4 (1975), pp. 643–663.

[111] Petter Holme and M. E. J. Newman. "Nonequilibrium phase transition in the coevolution of networks and opinions". In: *Physical Review E* 74.5 (2006). DOI: 10.1103/physreve.74.056108.

[112] Petter Holme and Jari Saramäki. "Temporal networks". In: *Physics reports* 519.3 (2012), pp. 97–125.

[113] L. Horstmeyer and C. Kuehn. "Adaptive voter model on simplicial complexes". In: *Physical Review E* 101.2 (2020).

[114] S. Huang, Bao-Xin Xiu, and Yanghe Feng. "Modeling and simulation research on propagation of Public Opinion". In: *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)* (2016), pp. 380–384.

[115] Robert Huckfeldt, Paul E Johnson, and John Sprague. *Political disagreement: The survival of diverse opinions within communication networks*. Cambridge University Press, 2004.

[116] Robert Huckfeldt, Jeanette Morehouse Mendez, and Tracy Osborn. "Disagreement, ambivalence, and engagement: The political consequences of heterogeneous networks". In: *Political Psychology* 25.1 (2004), pp. 65–95.

[117] Xie Hui et al. "Evolution of Bounded Confidence Opinion in Social Networks". In: (2017).

[118] David Hume. *A Treatise of Human Nature: being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects. Of the Understanding*. 1739.

[119] Iacopo Iacopini et al. "Simplicial models of social contagion". In: *Nature communications* 10.1 (2019), p. 2485.

[120] Gerardo Iñ iguez et al. "Opinion and community formation in coevolving networks". In: *Physical Review E* 80.6 (2009). DOI: 10.1103/physreve.80.066119.

[121] Roberto Interdonato et al. "Feature-rich networks: going beyond complex network topologies". In: *Applied Network Science* 4.1 (2019), pp. 1–13.

[122] S. Iyengar and Kyu S. Hahn. "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use". In: *Journal of Communication* 59 (2009), pp. 19–39.

[123] Dirk Jacobmeier. "Focusing of opinions in the Deffuant model: First impression counts". In: *International Journal of Modern Physics C* 17.12 (2006), pp. 1801–1808. DOI: 10.1142/S0129183106010108.

[124] Dirk Jacobmeier. "Multidimensional Consensus model on a Barabasi-Albert network". In: *International Journal of Modern Physics C* 16.04 (2005), pp. 633–646. DOI: 10.1142/S0129183105007388.

[125] Wander Jager and Frédéric Amblard. "Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change". In: *Computational & Mathematical Organization Theory* 10.4 (2005), pp. 295–303.

[126] Stephanie Jean Tsang. "Cognitive discrepancy, dissonance, and selective exposure". In: *Media Psychology* 22.3 (2019), pp. 394–417.

[127] Myeong Rye Jeong et al. "Feeling displeasure from online social media postings: A study using cognitive dissonance theory". In: *Comput. Hum. Behav.* 97 (2019), pp. 231–240.

[128] Shuyang Jiang and Hu Wang. "Group polarization based on agent emotional characteristics and credibility". In: *Complexity* 2021 (2021), pp. 1–12.

[129] Harang Ju et al. "The network structure of scientific revolutions". In: *arXiv preprint arXiv:2010.08381* (2020).

[130] Unchitta Kan, Michelle Feng, and Mason A Porter. "An adaptive bounded-confidence model of opinion dynamics on networks". In: *Journal of Complex Networks* 11.1 (2023), pp. 415–444.

[131] Brian Karrer and Mark EJ Newman. "Stochastic blockmodels and community structure in networks". In: *Physical review E* 83.1 (2011), p. 016107.

[132] Mikko Kivelä et al. "Multilayer networks". In: *Journal of Complex Networks* 2.3 (2014), pp. 203–271. DOI: 10.1093/comnet/cnu016.

[133] Joseph T Klapper. "The effects of mass communication." In: (1960).

[134] Aleksejus Kononovicius et al. "Empirical analysis and agent-based modeling of the Lithuanian parliamentary elections". In: *Complexity* 2017 (2017).

[135] B. Kozma and A. Barrat. "Consensus formation on coevolving networks: groups' formation and structure". In: *Journal of Physics A* 41 (2008), p. 224020.

[136] Emily Kubin and Christian von Sikorski. "The role of (social) media in political polarization: a systematic review". In: *Annals of the International Communication Association* 45.3 (2021), pp. 188–206.

[137] Bhushan Kulkarni et al. "SLANT+: A nonlinear model for opinion dynamics in social networks". In: *2017 IEEE International Conference on Data Mining (ICDM)* (2017), pp. 931–936.

[138] Z. Kunda. "The case for motivated reasoning". In: *Psychological bulletin* 108.3 (1990), p. 480.

[139] Evguenii Kurmyshev, Héctor A. Juárez, and Ricardo A. González-Silva. "Dynamics of bounded confidence opinion in heterogeneous social networks: Concord against partial antagonism". In: *Physica A: Statistical Mechanics and its Applications* 390.16 (2011), pp. 2945–2955. DOI: 10.1016/j.physa.2011.03.037.

[140] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. "Benchmark graphs for testing community detection algorithms". In: *Physical review E* 78.4 (2008), p. 046110. DOI: 10.1103/physreve.78.046110.

[141] Bibb Latané. "The psychology of social impact." In: *American psychologist* 36.4 (1981), p. 343.

[142] Richard R Lau et al. "Effect of media environment diversity and advertising tone on information search, selective exposure, and affective polarization". In: *Political Behavior* 39 (2017), pp. 231–255.

[143] E. Lawrence, John M. Sides, and Henry Farrell. "Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics". In: *Perspectives on Politics* 8 (2010), pp. 141–157.

[144] David Lazer et al. "Computational social science". In: *Science* 323.5915 (2009), pp. 721–723.

[145] Jonathan M Levine et al. "Beyond pairwise mechanisms of species coexistence in complex communities". In: *Nature* 546.7656 (2017), pp. 56–64.

[146] Grace J Li and Mason A Porter. "A bounded-confidence model of opinion dynamics with heterogeneous node-activity levels". In: *arXiv:2206.09490* (2022).

[147] Lin Li et al. "Consensus, Polarization and Clustering of Opinions in Social Networks". In: *IEEE Journal on Selected Areas in Communications* 31 (2013), pp. 1072–1083.

[148] Mingwu Li and Harry Dankowicz. "Impact of temporal network structures on the speed of consensus formation in opinion dynamics". In: *Physica A: Statistical Mechanics and its Applications* 523 (2019), pp. 1355–1370.

[149] Jan Lorenz. "Continuous Opinion Dynamics under Bounded Confidence: A Survey". In: *International Journal of Modern Physics C* 18 (2007), pp. 1819–1838.

[150] Jan Lorenz. "Heterogeneous bounds of confidence: meet, discuss and find consensus!" In: *Complexity* 15.4 (2010), pp. 43–52.

[151] Jan Lorenz and D. Urbig. "About the Power to Enforce and Prevent Consensus by Manipulating Communication Rules". In: *Adv. Complex Syst.* 10 (2007), pp. 251–269.

[152] Michael Maes and Lukas Bischofberger. "Will the Personalization of Online Social Networks Foster Opinion Polarization?" In: *Available at SSRN 2553436* 2553436 (2015).

[153] Gregory J Martin and Ali Yurukoglu. "Bias in cable news: Persuasion and polarization". In: *American Economic Review* 107.9 (2017), pp. 2565–2599.

[154] T Vaz Martins, Miguel Pineda, and Raul Toral. "Mass media and repulsive interactions in continuous-opinion dynamics". In: *Europhysics Letters* 91.4 (2010), p. 48003.

[155] Nolan McCarty. *Polarization: What everyone needs to know®*. Oxford University Press, 2019.

[156] Gary Mckeown and Noel Sheehy. "Mass media and polarisation processes in the bounded confidence model of opinion dynamics". In: *Journal of Artificial Societies and Social Simulation* 9.1 (2006).

[157] Miller McPherson, Lynn Smith-Lovin, and James M Cook. "Birds of a Feather: Homophily in Social Networks". In: *Annual Review of Sociology* 27.1 (2001), pp. 415–444. DOI: 10.1146/annurev.soc.27.1.415.

[158] X. Flora Meng, Robert A. Van Gorder, and Mason A. Porter. "Opinion formation and distribution in a bounded-confidence model on various networks". In: *Physical Review E* 97.2 (2018), p. 022312. DOI: 10.1103/PhysRevE.97.022312. URL: https://link.aps.org/doi/10.1103/PhysRevE.97.022312 (visited on 09/21/2023).

[159] Stanley Milgram. "The small world problem". In: *Psychology today* 2.1 (1967), pp. 60–67.

[160] John Stuart Mill. *M. de Tocqueville on Democracy in America*. Vol. 2. John W. Parker and son, 1859.

[161] James Moody. "Race, school integration, and friendship segregation in America". In: *American journal of Sociology* 107.3 (2001), pp. 679–716.

[162] A. J. Morales et al. "Twitter shows the two sides of venezuela". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25.3 (2015).

[163] Virginia Morini, Laura Pollacci, and Giulio Rossetti. "Capturing Political Polarization of Reddit Submissions in the Trump Era." In: *SEBD* (2020), pp. 80–87.

[164] Virginia Morini, Laura Pollacci, and Giulio Rossetti. "Toward a standard approach for echo chamber detection: reddit case study". In: *Applied Sciences* 11.12 (2021), p. 5390.

[165] Mehdi Moussaıd et al. "Social influence and the collective dynamics of opinion formation". In: *PloS one* 8.11 (2013).

[166] Diana C Mutz. "The consequences of cross-cutting networks for political participation". In: *American Journal of Political Science* (2002), pp. 838–855.

[167] Cecilia Nardini, Balázs Kozma, and Alain Barrat. "Who's Talking First? Consensus or Lack Thereof in Coevolving Opinion Formation Models". In: *Physical Review Letters* 100.15 (2008), p. 158701. DOI: 10.1103/PhysRevLett.100.158701.

[168] Leonie Neuhäuser et al. "Opinion dynamics with multi-body interactions". In: *International Conference on Network Games, Control and Optimization* (2021), pp. 261–271.

[169] Mark EJ Newman. "The structure and function of complex networks". In: *SIAM review* 45.2 (2003), pp. 167–256.

[170] Raymond S Nickerson. "Confirmation bias: A ubiquitous phenomenon in many guises". In: *Review of general psychology* 2.2 (1998), pp. 175–220.

[171] Hossein Noorazar. "Recent advances in opinion propagation dynamics: a 2020 survey". In: *The European Physical Journal Plus* 135.6 (2020). DOI: 10.1140/epjp/s13360-020-00541-2.

[172] Hossein Noorazar et al. "From classical to modern opinion dynamics". In: *International Journal of Modern Physics C* 31.07 (2020), p. 2050101. DOI: 10.1142/s0129183120501016.

[173] Maria Nordbrandt. "Affective polarization in the digital age: Testing the direction of the relationship between social media and users' feelings for outgroup parties". In: *New media & society* (2021).

[174] Martin A Nowak and David C Krakauer. "The evolution of language". In: *Proceedings of the National Academy of Sciences* 96.14 (1999), pp. 8028–8033.

[175] Brendan Nyhan and Jason Reifler. "When Corrections Fail: The Persistence of Political Misperceptions". In: *Political Behavior* 32 (2010), pp. 303–330.

[176] V. Pansanella, G. Rossetti, and L. Milli. "Modeling algorithmic bias: simplicial complexes and evolving network topologies." In: *Appl Netw Sci* (2022). DOI: https://doi.org/10.1007/s41109-022-00495-7.

[177] Valentina Pansanella, Giulio Rossetti, and Letizia Milli. "From Mean-Field to Complex Topologies: Network Effects on the Algorithmic Bias Model". In: Springer International Publishing, 2022, pp. 329–340.

[178] Valentina Pansanella et al. "Change my Mind: Data Driven Estimate of Open-Mindedness from Political Discussions". In: *Complex Networks and Their Applications XI: Proceedings of The Eleventh International Conference on Complex Networks and Their Applications: COMPLEX NETWORKS 2022–Volume 1* (2023), pp. 86–97.

[179] Valentina Pansanella et al. "Mass Media Impact on Opinion Evolution in Biased Digital Environments: a Bounded Confidence Model". In: *Scientific Reports* (2023).

[180] Christophe Paris et al. "COVID-19 vaccine hesitancy among healthcare workers". In: *Infectious diseases now* 51.5 (2021), pp. 484–487.

[181] Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

[182] John V Pavlik. *Media in the digital age*. Columbia University Press, 2008.

[183] Antonio F Peralta, János Kertész, and Gerardo Iñiguez. "Opinion dynamics in social networks: From models to data". In: *arXiv preprint arXiv:2201.01322* (2022).

[184] Antonio F Peralta, János Kertész, and Gerardo Iñiguez. "Opinion formation on social networks with algorithmic bias: dynamics and bias imbalance". In: *Journal of Physics: Complexity* 2.4 (2021).

[185] Antonio F Peralta et al. "Effect of algorithmic bias and network structure on coexistence, consensus, and polarization of opinions". In: *Physical Review E* 104.4 (2021), p. 044312.

[186] N. Perra and L. E. C. Rocha. "Modelling opinion dynamics in the age of algorithmic personalisation". In: *Scientific Reports* (2019).

[187] Giovanni Petri and Alain Barrat. "Simplicial Activity Driven Model". In: *Phys. Rev. Lett.* 121 (22 2018), p. 228301. DOI: 10.1103/PhysRevLett.121.228301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.228301.

[188] M. Pineda and G. M. Buendía. "Mass media and heterogeneous bounds of confidence in continuous opinion dynamics". In: *Physica A-statistical Mechanics and Its Applications* 420 (2015), pp. 73–84.

[189] Walter Quattrociocchi, R. Conte, and E. Lodi. "Opinions Manipulation: Media, Power and Gossip". In: *Adv. Complex Syst.* 14 (2011), pp. 567–586.

[190] José J Ramasco, Sergey N Dorogovtsev, and Romualdo Pastor-Satorras. "Self-organization of collaboration networks". In: *Physical review E* 70.3 (2004), p. 036106.

[191] Manoel Horta Ribeiro, Veniamin Veselovsky, and Robert West. "The Amplification Paradox in Recommender Systems". In: *Proceedings of the International AAAI Conference on Web and Social Media* (2023).

[192] Giulio Rossetti et al. "NDlib: a python library to model and analyze diffusion processes over complex networks". In: *International Journal of Data Science and Analytics* 5.1 (2018), pp. 61–79.

[193] Rohit Sahasrabuddhe, Leonie Neuhäuser, and Renaud Lambiotte. "Modelling non-linear consensus dynamics on hypergraphs". In: *Journal of Physics: Complexity* 2 (2020). URL: https : / / api . semanticscholar . org / CorpusID : 220646971.

[194] Kazutoshi Sasahara et al. "Social influence and unfollowing accelerate the emergence of echo chambers". In: *Journal of Computational Social Science* 4 (2019), pp. 381–402. URL: https : / / api . semanticscholar . org / CorpusID : 221297607.

[195] Hendrik Schawe and Laura Hern'andez. "Higher order interactions destroy phase transitions in Deffuant opinion dynamics model". In: *Communications Physics* 5 (2021). URL: https://api.semanticscholar.org/CorpusID:244527584.

[196] Hendrik Schawe and Laura Hernández. "When open mindedness hinders consensus". In: *Scientific reports* 10.1 (2020), p. 8273.

[197] Thomas C Schelling. "Dynamic models of segregation". In: *Journal of mathematical sociology* 1.2 (1971), pp. 143–186.

[198] F. Schweitzer. "Sociophysics". In: *Phys Today* (2018).

[199] Yilun Shang. "An agent based model for opinion dynamics with random confidence threshold". In: *Communications in Nonlinear Science and Numerical Simulation* 19.10 (2014), pp. 3766–3777.

[200] Ali M Shropshire, Renee Brent-Hotchkiss, and Urkovia K Andrews. "Mass media campaign impacts influenza vaccine obtainment of university students". In: *Journal of American College Health* 61.8 (2013), pp. 435–443.

[201] Wesley Shrum, Neil H Cheek Jr, and Saundra MacD. "Friendship in school: Gender and racial homophily". In: *Sociology of Education* (1988), pp. 227–239.

[202] Doron Shultziner and Yelena Stukalin. "Distorting the news? The mechanisms of partisan media bias and its effects on news production". In: *Political Behavior* 43.1 (2021), pp. 201–222.

[203] Omid Askari Sichani and Mahdi Jalili. "Inference of hidden social power through opinion formation in complex networks". In: *IEEE Transactions on Network Science and Engineering* 4.3 (2017), pp. 154–164.

[204] A. Sîrbu et al. "Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model". In: *PLoS ONE* 14 (2019).

[205] Alina Sırbu et al. "Opinion dynamics: models, extensions and external effects". In: Springer, 2017, pp. 363–401. ISBN: 9783319256580.

[206] Jeffrey A Smith, Miller McPherson, and Lynn Smith-Lovin. "Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004". In: *American Sociological Review* 79.3 (2014), pp. 432–456.

[207] Robert Snow and D Altheide. "Media Logic". In: *Beverly Hills* 8 (1979), pp. 1094–1096.

[208] P. Sobkowicz. "Quantitative Agent Based Model of Opinion Dynamics: Polish Elections of 2015". In: *PLoS ONE* 11 (2016).

[209] Paweł Sobkowicz. "Social Depolarization and Diversity of Opinions—Unified ABM Framework". In: *Entropy* 25.4 (2023), p. 568.

[210] D. Stauffer and H. Meyer-Ortmanns. "SIMULATION OF CONSENSUS MODEL OF DEFFUANT et al. ON A BARABÁSI–ALBERT NETWORK". In: *International Journal of Modern Physics C* 15 (2004), pp. 241–246.

[211] Dietrich Stauffer et al. "Discretized Opinion Dynamics of The Deffuant Model on Scale-Free Networks". In: *Journal of Artificial Societies and Social Simulation* 7.3 (2004), pp. 1–7.

[212] Alexander J. Stewart et al. "Information gerrymandering and undemocratic decisions". In: *Nature* 573 (2019), pp. 117–121.

[213] Jessica Su, Aneesh Sharma, and Sharad Goel. "The Effect of Recommendations on Network Structure". In: *Proceedings of the 25th International Conference on World Wide Web*. WWW '16 (2016), pp. 1157–1167. DOI: 10.1145/2872427.2883040. URL: https://doi.org/10.1145/2872427.2883040.

[214] Cass R. Sunstein. *Republic.Com 2.0*. USA: Princeton University Press, 2007. ISBN: 0691133565.

[215] K. Sznajd-Weron and J. Sznajd. "Opinion evolution in closed community". In: *HSC Research Reports* (2000).

[216] Taro Takaguchi and Naoki Masuda. "Voter model with non-Poissonian interevent intervals". In: *Phys. Rev. E* 84 (3 2011), p. 036115. DOI: 10.1103/PhysRevE.84.036115. URL: https://link.aps.org/doi/10.1103/PhysRevE.84.036115.

[217] J. Timothy. "How does propaganda influence the opinion dynamics of a population?" In: *ArXiv* abs/1703.10138 (2017).

[218] Francisco Nataniel Macedo Uchôa et al. "Influence of the mass media and body dissatisfaction on the risk in adolescents of developing eating disorders". In: *International journal of environmental research and public health* 16.9 (2019), p. 1508.

[219] Adriano Udani, David C Kimball, and Brian Fogarty. "How local media coverage of voter fraud influences partisan perceptions in the United States". In: *State Politics & Policy Quarterly* 18.2 (2018), pp. 193–210.

[220] C. M. Valensise, M. Cinelli, and W. Quattrociocchi. *The dynamics of online polarization*. 2022.

[221] Sebastián Valenzuela et al. "The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation". In: *Digital Journalism* 7.6 (2019), pp. 802–823.

[222] Antoine Vendeville, Benjamin Guedj, and Shi Zhou. "Forecasting elections results via the voter model with stubborn nodes". In: *Applied Network Science* 6 (2021), pp. 1–13.

[223] S. Verducci. "Critical thinking and and open-mindedness in polarized times". In: *Encounters in Theory and History of Education* 20.1 (2019), pp. 6–23.

[224] Michela Del Vicario et al. "Modeling confirmation bias and polarization". In: *Scientific Reports* 7 (2016). URL: https://api.semanticscholar.org/CorpusID:11189837.

[225] K Viswanath, Shoba Ramanadhan, and Emily Z Kontos. "Mass media". In: Springer, 2007, pp. 275–294.

[226] Duncan J Watts and Steven H Strogatz. "Collective dynamics of 'small-world'networks". In: *nature* 393.6684 (1998), pp. 440–442. DOI: https://doi.org/10.1038/30918.

[227] J. Weatherall, Cailin O'Connor, and Justin P. Bruner. "How to Beat Science and Influence People: Policymakers and Propaganda in Epistemic Networks". In: *The British Journal for the Philosophy of Science* 71 (2020), pp. 1157–1186.

[228] G. Weisbuch. "Bounded confidence and social networks". In: *The European Physical Journal B* 38 (2004), pp. 339–343.

[229] Christine B Williams. "Introduction: Social media, political marketing and the 2016 US election". In: Routledge, 2018, pp. 1–5.

[230] Magdalena Wojcieszak and Benjamin R Warner. "Can interparty contact reduce affective polarization? A systematic test of different forms of intergroup contact". In: *Political Communication* 37.6 (2020), pp. 789–811.

[231] F. Xiong and Yun Liu. "Opinion formation on social media: an empirical approach." In: *Chaos* 24 1.1 (2014), p. 013130.

[232] Ihsan Yilmaz and Galib Bashirov. "Religious homophily and friendship: Socialisation between Muslim minority and Anglo majority youth in Australia". In: *Journal of Youth Studies* 26.6 (2023), pp. 768–785.

[233] Damián H. Zanette and Santiago Torres Gil. "Opinion spreading and agent segregation on evolving networks". In: *Physica D: Nonlinear Phenomena* 224 (2006), pp. 156–165. URL: https://api.semanticscholar.org/CorpusID:119770502.

[234] Quanbo Zha et al. "Opinion dynamics in finance and business: a literature review and research opportunities". In: *Financial Innovation* 6 (2021), pp. 1–22.