IMT School for Advanced Studies, Lucca

Lucca, Italy

The Role of Confirmation Bias in the Emergence of Echo Chambers: A Data-Driven Approach

PhD Program in Computer, Decision, and Systems Science

XXIX Cycle

By

Michela Del Vicario

2016

The dissertation of Michela Del Vicario is approved.

Program Coordinator: Prof. Rocco De Nicola, IMT School for Advanced Studies, Lucca

Supervisor: Prof. Guido Caldarelli, IMT School for Advanced Studies, Lucca

Cosupervisor: Dr. Walter Quattrociocchi, IMT School for Advanced Studies, Lucca

The dissertation of Michela Del Vicario has been reviewed by:

Prof. Sabrina Gaito, University of Milan

Prof. Andrea Gabrielli, University of Rome, *La Sapienza*

IMT School for Advanced Studies, Lucca

2016

Contents

Acknowledgements **vii**

Vita and Publications **viii**

Abstract **xi**

1 Introduction **1**

2 Background and Literature Review **6**

3 The Spreading of Misinformation Online: *A Data-Driven Model of Echo Chambers Formation* **14**

3.1 Materials and Methods 15

3.1.1 Data Description 15

3.1.2 Basic Definitions 16

3.2 Analysis of Observable Data 18

3.3 The Model 26

3.4 Simulation Results 29

3.5 Concluding Remarks 39

4 Echo Chambers: Emotional Contagion and Group Polarization on Facebook **40**

4.1 Materials and Methods 42

4.2 Community Evolution 46

4.3 Users' Sentiment Analysis 56

4.4 Evolution of the Sentiment inside the Communities 63

4.5 Concluding Remarks 67

5 Modeling Opinion Dynamics on Networks: *The role of Confirmation Bias and Polarization in Opinions Formation* **68**
5.1 The Bounded Confidence Model (BCM) 71
5.2 Results and Discussion 73
5.2.1 Models 73
5.2.2 Simulation Results 74
5.2.3 Final Distribution of Peaks 78
5.2.4 Mean Field Approximation 82
5.2.5 Simulation Results for RBCM 84
5.2.6 Simulation Results for ER and SW 86
5.3 Concluding Remarks 90

6 Conclusions and Future Works **91**

A Other Works **94**
A.1 Social Determinants of Content Selection in the Age of (Mis) Information 95
A.2 Homophily and Polarization in the Age of Misinformation 96
A.3 Modeling Networks with a Growing Feature-Structure . . 98
A.4 Trend of Narratives in the Age of Misinformation 98
A.5 Emotional Dynamics in the Age of Misinformation 99
A.6 Debunking in a World of Tribes 100
A.7 User Polarization on Facebook and Youtube 102

References **105**

Acknowledgements

The thesis is based on co-authored published or about to be published articles:

1. Chapter 3 is the reproduction of the paper *"The Spreading of Misinformation online"*(29), published in the proceeding of the National Academy of Science and co-authored with Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi.
2. Chapter 4 is the reproduction of the paper *"Echo Chambers: Emotional Contagion and Group Polarization"*(30), under review and available as a preprint at *arXiv:1607.01032*. Tis paper is co-authored with Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi.
3. Chapter 5 is the reproduction of the paper *"Modeling Confirmation Bias and Polarization"*, under review and available as a preprint at *arXiv:1509.00189*(31). This paper is co-authored with Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi.

Vita

December 20, 1988	Born, Sora (FR), Italy
2010	Bachelor Degree in Mathematics Final mark: 100/110 University of Rome *La Sapienza*
2013	Mater Degree in Mathematics Final mark: 110/110 University of Rome *La Sapienza*
2016	Visiting Postgraduate Scholar Center for Polymer Studies - Boston University Boston, United States
Currently	PhD Program IMT - School for Advanced Studies Lucca Curriculum: Computer, Decision, and System Science. Fellow member of CSSLab Computational Social Science Laboratory Research Unit: Networks.

Publications

1. Alessandro Bessi, Guido Caldarelli, **Michela Del Vicario**, Antonio Scala and Walter Quattrociocchi, "Social determinants of content selection in the age of (mis)information", *Proceedings of SocInfo 2014*, 2014.
2. Alessandro Bessi, Fabio Petroni, **Michela Del Vicario**, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, Walter Quattrociocchi, "Viral misinformation: the role of homophily and polarization", *Proceedings of WWW'15, WebSci Track Papers & Posters*, 2015.
3. Alessandro Bessi, Fabiana Zollo, **Michela Del Vicario**, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi, "Trend of Narratives in the Age of Misinformation", *PLoS ONE*, 2015.
4. Fabiana Zollo, Petra K. Novak, **Michela Del Vicario**, Alessandro Bessi, Igor Mozetic, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi, "Emotional Dynamics in the Age of Misinformation", *PLoS ONE*, 2015.
5. Irene Crimaldi, **Michela Del Vicario**, Greg Morrison, Walter Quattrociocchi, Massimo Riccaboni, "Modeling Networks with a Growing Feature-Structure", *arXiv preprint*, 2015.
6. **Michela Del Vicario**, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, Walter Quattrociocchi, "The Spreading of Misinformation Online", *Proceedings of the National Academy of Science*, 2016.
7. Fabiana Zollo, Alessandro Bessi, **Michela Del Vicario**, Antonio Scala, Guido Caldarelli, Luis Shekhtman, Shlomo Havlin, Walter Quattrociocchi, "Debunking in a World of Tribes", *arXiv preprint*, 2015.
8. Alessandro Bessi, Fabiana Zollo, **Michela Del Vicario**, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi, "Users Polarization on Facebook and Youtube", *(to appear on) PLoS ONE*, 2016.
9. Alessandro Bessi, Fabio Petroni, **Michela Del Vicario**, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, Walter Quattrociocchi, "Homophily and polarization in the age of misinformation", *(to appear on) Eur. Phys. J. S.T.*, 2016.
10. **Michela Del Vicario**, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, Walter Quattrociocchi, "Echo Chambers: Emotional Contagion and Group Polarization on Facebook", *arXiv preprint* (under review), 2016.

11. **Michela Del Vicario**, Antonio Scala, Guido Caldarelli, H Eugene Stanley, Walter Quattrociocchi, "Modeling confirmation bias and polarization", *arXiv preprint* (under review), 2016.

12. **Michela Del Vicario**, Alessandro Bessi, Guido Caldarelli, Fabiana Zollo, "Structural Patterns of the Occupy Movement on Facebook", *Proceedings of Complex Networks* (accepted), 2016.

Presentations

1. Alessandro Bessi, Guido Caldarelli, **Michela Del Vicario**, Antonio Scala and Walter Quattrociocchi, "Social determinants of content selection in the age of (mis)information", at *SocInfo 2014*, 2014.

2. **Michela Del Vicario**, "An analysis of the Spreading of Misinformation Online", at *SoBigData Workshop, TU Delft*, 2016.

3. **Michela Del Vicario**, "The Spreading of Misinformation Online", at *Boston University, Center for Polymer Studies*, 2016.

4. **Michela Del Vicario**, "Modeling the Emergence of Two Opinions Under Confirmation Bias", at *Boston University, Center for Polymer Studies*, 2016.

Abstract

The thesis focuses on the modeling of information diffusion and opinion dynamics on online social networks, using a data-driven approach accounting for the presence of social contagion and cognitive effects such as confirmation bias, cognitive dissonance, and backfire effect.

We a) analyze data from a sample of 1.2M of users on the Italian Facebook pages, focused of scientific and conspiracy contents, by means of quantitative methods, statistical analysis, and sentiment analysis; b) develop a data-driven percolation model of signed edges to mimic the information spreading and a theoretical model of opinions formation.

We provide evidences that the diffusion of information, either substantiated or not, is promoted by confirmation bias and homophily. This process in turn generates and fosters the formation of homogeneous polarized clusters, the echo chambers. Users' emotional behavior seems to be affected by their engagement within the community. An higher involvement in the echo chamber, resolves in a more negative emotional state.

Lastly, we develop a model of opinions formation that takes into account both confirmation bias and social influence as triggering factors for the group polarization on social networks. Our model is able to reproduce the dynamics we observed on Facebook.

Chapter 1

Introduction

People perceptions, knowledge, beliefs, and opinions about the world and its evolution get (in)formed and modulated through the information they can access, most of which coming from newspapers, television (18), and, more recently, the Internet. However, the web, social networks, and micro-blogging platforms have changed the way we can produce and consume information. In particular, large social networks, with their user-provided contents, facilitate the study of social phenomena related to the emergence, production, and consumption of information (16; 19). The relationship between the information and its effect on individual opinions is central if we consider that on the web users tend to consume and modulate their opinion upon unverified, unsubstantiated (and even false) information in the same way (20; 21; 22).

The diffusion of social media caused a shift of paradigm in the creation and consumption of information. We passed from a mediated (e.g., by journalists) to a more disintermediated selection process. Such a disintermediation elicits the tendencies of the users to first select information adhering to their system of beliefs – i.e., confirmation bias – and second, form groups of like-minded people where they polarize their opinion – i.e. echo chamber.

One of the most analyzed social phenomena is the *social contagion* (11; 17), that may be defined as the set of social phenomena that can and

do spread though social networks. The study of the portion of social contagion that is observed on *Online Social Networks (OSN)* is taking a great advantage from the shift to data-driven models, as the accessibility to OSN data makes it possible to observe social phenomena on millions of individuals. Recent approaches to the study of social contagion and other relevant social topics such as *information diffusion*, or *communities evolution dynamics*, have been carried on both by theoretical models (3; 4; 5; 6; 7; 8) and data-driven analysis (9; 10; 11; 12; 13; 14; 15; 16). The models used to describe such diverse social phenomena can be either *explanatory*, aiming at inferring the underlying process given a complete realization of it, or *predictive* with the final purpose of anticipate the process by learning from past realizations of it. Theoretical models usually assume simple dynamical rules for the description of micro-level interactions, paying the cost of an unavoidable reductionism in favor of simplicity of representation. This problem can be overcome by developing data-driven models as a combination of the increasingly interdisciplinary collaborations and the unprecedented availability of detailed online data. Interesting and sometimes unexpected behaviors emerge from this new modeling approach, opening the challenge towards their theoretical understanding.

There are various forces that lead to social contagion and make products, rumors, and social movements spread from person to person. All these mechanisms can be treated as *influence*. In addition, there are some other mechanisms such as *homophily*[1] and *environmental factors*, that create effects apparently resembling the diffusion of social contagion, but having distinct underlying causes. Therefore, distinguishing influence from homophily or environmental factors is of primary importance when talking about social contagion.

Among the most common cognitive phenomena influencing the opinion formation and diffusion we recall the *confirmation bias*, *cognitive dissonance*, and *backfire effect* (23; 24; 25). Confirmation bias is defined as the tendency to acquire or process new information in a way that confirms

[1]Homophily is defined as the tendency of one individual to establish social relationships with individuals similar to herself.

one's preconceptions and avoids contradiction with prior beliefs (23), and it usually refers to an unconscious act of selection. Cognitive dissonance (24) was first theorized by Festinger in the mid-50s and refers to a state of discomfort generated by the coexistence of conflicting beliefs about one cognition. He defined this unpleasant state as dissonance and theorized its degree in relation to a cognition to be $D/(D+C)$, where D is the sum of dissonant beliefs to a particular cognition and C is the sum of consonant ones. Conversely, the backfire effect (25) refers to the empirical observation that in some particular cases the response to persuasion is to move further distant to the persuader's opinion. It is important to take into account such cognitive mechanisms, together with the social influence and homophily, when studying the diffusion of information and opinions since it is widely acknowledged that people shape their opinions on the basis of both internal and external forces (26).

The diffusion of information and rumors has been enhanced by the structure of the web and by the disintermediation that it promotes. On the Internet every user can produce and easily access all the online knowledge and news instantly and without any intermediaries. Specifically, the quality of the online information could be damaged by this environment while at the same time the spreading of *digital misinformation* could be promoted. Digital misinformation refers to the diffusion of false or misleading information online. The risk of misinformation spreading have been underlined by the global risk reports of the World Economic Forum (WEF) (27; 28). Characterizing the diffusion of information online and predicting the potential virality[2] of either official news or rumors, is becoming a particularly compelling problem.

Given an online social system, it can be modeled as a graph where each individual is a vertex and an edge exists between two vertices if there is a social relationship between the corresponding individuals in the OSN. When dealing with social phenomena, it is important to take into account the effect of social and psychological mechanisms trigger-

[2]Virality is here intended as the rapid and wide diffusion of a phenomenon, e.g., online information is considered to be viral if consumed and shared by millions of users in a relatively short time.

ing the formation of opinions and beliefs. Moreover, the collection of observables is a crucial preliminary step: given a phenomenon it is necessary to get the right observables that will characterize it properly. In order to build models suited for the explanation and prediction of processes on social systems, it is necessary to take into account social traces and their temporal order. The relationship between the individuals and the information sources should also be considered, e.g., how information is produced, how people get informed, how information sources interact among themselves (in terms of content selection). In addition, we should always bear in mind that individuals' actions are not necessarily an expression of linear relation between what is rational and their own opinions.

Thesis Advances

The present thesis is mainly concerned in the modeling of information diffusion and opinion dynamics on OSN, using a data-driven approach accounting for the presence of social contagion and cognitive effects such as confirmation bias, cognitive dissonance, and backfire effect. Also, we aim to uncover the underlying process of digital misinformation spreading by building models that can mimic real world phenomena starting from observable data.

The works collected in this thesis address different aspects of the online social dynamics, from the spreading of misinformation to the emergence of echo chambers and group polarization. We analyze data from a sample of 1.2M of users on the Italian Facebook pages focused of scientific and conspiracy-like contents.

We provide, through quantitative methods, evidences that the diffusion of information, either substantiated or not, is promoted by confirmation bias and homophily. This process in turn generates and fosters the formation of homogeneous polarized clusters, the echo chambers. We also reproduce this mechanism by means of a data-driven percolation model that accounts for the observed polarization and homophily.

Once observed the existence of group polarization on our sample

data, we analyze the community behavior by accounting for the engagement and the emotional dynamics of users. Users' emotional behavior seems to be affected by their engagement within the community. An higher involvement in the echo chamber, resolves in a more negative emotional state. Such a phenomenon appears in both users categories.

Lastly, we develop a model of opinion formation that takes into account both confirmation bias and social influence as triggering factors for the group polarization on social networks. Our model is able to reproduce the dynamics we observed on Facebook.

The thesis is structured as follows. In Chapter 2 we provide a review of the literature in the filed of modeling and data-driven analysis of social dynamics, in particular those related to information diffusion and digital misinformation. In Chapter 3 we report results already published in (29) on the spreading of misinformation. In Chapter 4 we analyze the community evolution and the emotional contagion relatively to two groups of Facebook users. In Chapter 5 we develop two new opinion dynamics models and provide simulation results and a mean field approximation. All the results in Chapters 4 and 5, come from the submitted papers (30; 31). Lastly in the Appendix A, results from other works are sketched.

Chapter 2

Background and Literature Review

The thesis focuses on the modeling of information diffusion and opinion dynamics on OSN, using a data-driven approach accounting for the presence of social contagion and cognitive effects such as confirmation bias, cognitive dissonance, and backfire effect. In this chapter we first introduce the idea of modeling social systems and present a review of the literature in the field. Then we give an overview of the literature in the field of data-driven analysis, with a specific attention to research works concerning information diffusion and digital misinformation.

Modeling Social Systems

Modeling social systems and the dynamics of opinion on them involves two levels of difficulties: inference and consistence. The former is the possibility to infer macroscopic phenomenology out of the microscopic dynamics. The latter is the possibility of defining families of models that capture the right aspects of phenomena, while being realistic at the same time. Social dynamics can not be modeled by means of the interaction of a large number of simple elements, as in the atoms and molecules case. A human behavior is already the complex outcome of many physiological

and psychological processes. Data-driven models offer the possibility to compare theoretical results with empirical ones, and hence determinate whether the trends seen in real data are compatible with models. *Agent-based models* represent another fundamental family of models in social dynamics. They are classes of computational devices able to simulate in parallel the interactions of many agents resembling real world phenomena.

The first attempts to model macroscopic social phenomena and opinion dynamics started off with simple assumptions and rules for the dynamics (32; 33). Early studies were generally carried on by social scientists, while in recent years the field also attracted the interest of statistical physicists (3). The simpler and more general class of models is represented by *behavioral models* where the attributes of agents are binary variables, as in the case of the voter model, the majority-rule model, and the Sznajd model (34; 35; 36; 37; 38; 39).

In the **Voter Model** (34; 35) there is a set of N agents each of which has a binary variable $s = \pm 1$ assigned at the beginning of the process. At each step a random agent $i \in \{1, \ldots, N\}$ is picked together with one of its neighbors $j \in Neigh(i)$, s_i will then be put equal to s_j, meaning that the agents imitate their neighbors. It is shown that starting from a disorder condition the voter model tends to an order state. The main question is whether full consensus is reached in a system of infinite size. Authors in (36) considered a d-dimensional hypercubic lattice and showed that, for infinite systems, consensus is reached only if $d \leq 2$, while consensus is invariably reached asymptotically if the system is finite (for any d), with convergence time T_N depending on the size N of the system.

The **Majority Rule Model** (37; 38) assumes the same framework of the voter model, N agents provided with a binary opinion function, with the difference that at the beginning there is a fixed fraction p_+ (p_-) of agent with opinion $+1$ (-1). At each step a group of r agents is selected at random and all the agents switch to the majority's opinion, with r depending on a given distribution. If r is even, a tie may happen, in this case one of the two opinions is chosen to prevail a priori. Let p_+^0 be the initial fraction of agents with opinion $+1$, it has been shown that exists a

threshold p_c such that, if $p_+^0 > p_c$ ($p_+^0 < p_c$), all agents will have opinion $+1$ (-1) in the long run, the time to reach consensus scales as $\log N$.

Along the path opened by Axelrod (33), new models in which opinions or cultures are represented by vectors of cultural traits have introduced the notion of bounded confidence: an agent will not interact with any other agent unless their opinions are close enough. Axelrod's model played a central role in the investigation of cultural dynamics. The **Dissemination of Culture Model** may be seen as a vectorial generalization of previous opinion dynamics models:

- Individuals are located on the vertices of a square grid each of them having an initial *vector of culture* made up of F features each of which can assume q traits. The culture is defined by Axelrod as a set of "beliefs, attitudes, and behaviors".

- An active site is picked at random along with one of its neighbors: with probability proportional to their similarity (the number of cultural features in common) they interact; as a result, one of the features on which they differ is chosen at random and the corresponding feature of the active site is put equal to that of its neighbor, hence augmenting their similarity and their chance to further interact in the future.

Two really important mechanisms are embodied in this model: social influence, in changing the active site cultural traits, and homophily, two sites only interact if they are similar to each other. While the combination of these two aspects is generally regarded as a source of social reinforcement by social scientists, it is shown that under certain circumstances a *persistence of diversity* appears (33).

In other instances, further realism has been introduced by the use of continuous opinion variables. In the **Bounded Confidence Model** (5; 40) the initial state includes a population of N agents, arranged on a complex network G, each of which holds an initial opinion x_i, $i \in \{1, \ldots, N\}$ uniformly distributed in the interval $[0, 1]$. At each time step a user i is picked at random, together with one of its neighbors (also at random), if the difference between their opinions (at that specific time step t) exceeds

a certain threshold ε nothing happens, otherwise, if $|x_i(t) - x_j(t)| < \varepsilon$, they readjust their opinions according to the following rule:

$$\begin{cases} x_i(t+1) = x_i(t) + \mu[x_j(t) - x_i(t)] \\ x_j(t+1) = x_j(t) + \mu[x_i(t) - x_j(t)], \end{cases}$$

where μ is the convergence rate, generally taken in the interval $[0, 1/2]$ for the simulations. Its evolution was initially studied through Monte Carlo simulations (5). The number of peaks of opinions depends on the value of the threshold ε, while μ and N only account for the speed of convergence.

Successive studies (41; 42) analytically showed that for ε big enough consensus is reached. When $\varepsilon \geq 1$ the final state is characterized by a single opinion located in the middle of the interval and, as long as $\varepsilon \geq 1/2$, this situation persists. For smaller values of the threshold ε, it has been shown by numerical simulations that consensus is not reached and the opinion evolves into clusters that are separated by a distance larger than ε. Once each cluster is isolated it evolves into a Dirac delta function as in the case $\varepsilon \geq 1$. The final distribution consists of a series of non interacting clusters at locations x_i with masses m_i:

$$\mathbb{P}_\infty(x) = \sum_{i=1}^{r} m_i \delta(x - x_i),$$

where r is the number of evolving opinion clusters (41). All different clusters $i \neq j$ must fulfill the condition $|x_i - x_j| > \varepsilon$.

A model that integrates social influence and homophily is that proposed by Holme and Newmann (43). The authors developed a simple model that considers a set of N nodes and M edges, where each node i holds one of G possible opinions on a certain topic, g_i. Both the M edges and the opinions are initially distributed uniformly at random. Then at each step a vertex i is chosen at random and if its degree is zero nothing happens, otherwise with probability ϕ one of i's edges is picked at random, the link is broken, and rewired with one of the nodes holding the same opinion of i's. While, with probability $1 - \phi$, a random neighbor j (of i's) is picked and g_i is set equal to g_j. The model segregates into a

set of communities such that no individual has any acquaintances with whom they disagree.

A different line of studies (8; 44) proposed a **Non Consensus Opinion Model** that allowed for the stable coexistence of two opinions by also considering the opinion of the user herself when applying the majority rule update (8), while in (44) the competition between two groups is investigated by the introduction of a set of contrarians in one of the two. The survival of a two-opinions state is studied in (14) from a different point of view, considering the emergence of spontaneous recovery of failed nodes and the majority rule update. Both these models assume only two opinion states (± 1) and a majority rule update, with the novelty of accounting for the individual opinion (8; 44) and for an external source of influence (14).

Other models (26; 45) explored the opinion dynamics under cognitive phenomena, e.g., confirmation bias, cognitive dissonance, and backfire effect. All these models are based on simple assumptions that are often far from reality. However, the availability of OSN data offers a new approach taht makes it possible to explain phenomena starting from data.

Data-Driven Analysis

Information Diffusion

Recent studies explored the evolution of social phenomena on OSN from observable data. One of the most investigated aspects is that of **structural properties** affecting social behavior (10; 11; 46).

Weak ties proved to be fundamental for the information diffusion. However, in (46), the strength of weak ties is debated arguing that, while long ties are helpful in the spreading of innovation and social movements, they are not enough for spreading the social reinforcement necessary to act on that information. In (11) the effect of network structure on users' personal decision is studied within an artificial social framework built for the experiment. Authors found evidence of an effect of reinforcement in the adoption of the investigated behavior for clustered net-

works, where many redundant ties are present. A more recent work (10), again concerned on the effect of network topology on individuals' decisions, analyzes the response to friends' invitations to join Facebook and the subsequent engagement in the platform. One of the main results is that a user is more likely to join Facebook if the friends inviting her form a large number of connected components. The number of different connected components, representing separate social contexts for the user, proved to be more influential than the number of friends itself.

The **influential hypothesis** (47; 48; 49) was first formalized through the *two-step flow* (47) and it is based on the idea that a minority of individuals is able to influence an exceptional number of their peers. In the model a small minority of "opinion leaders" acts as intermediary between the mass media and the majority of society. In (49) authors challenge the influential hypothesis role in diffusion processes. They performed a series of simulations on diffusion models: influencers appear to be only modestly more important than average individuals. Most social change is driven by *easily influenced individuals influencing other easily influenced individuals* (49).

Another interesting problem is the possibility to **distinguish influence from homophily** (50; 51; 52). In (50), authors studied the global network of instant messaging traffic among about 30 millions users on *Yahoo.com*, with complete data on the adoption of a mobile service application and precise dynamic behavioral data on users. The effect of homophily proved to explain more than 50% of the phenomena perceived as contagion. The methodologies used to isolate influence and homophily are based on a dynamic matched sample of treated and untreated users, where a user is treated if exposed to the news of adoption of the application by friends. In (51) the role of social network and exposure to friends' activities in information re-sharing on Facebook is analyzed. Through controlled experiments in which the users were divided into exposed and not exposed to friends' re-sharing, authors were able to isolate contagion from other confounding effects like homophily. They claimed that in the exposed case there is a considerably higher chance to share contents. Moreover, a surprising result is that new information of-

ten spreads through weak ties, although stronger ties are individually more influential.

A fundamental class of observable phenomena is that of **Cascades**[1] (15; 16; 53; 54), taking place in many environments: cultural fads, collective actions, diffusion of informations, norms, and innovations. Cascades represent rare phenomena and it is necessary to be careful in order to avoid any kind of bias in the sample, – e.g., only considering large cascades.

In (53) a binary-decision model is specified for the prediction of real system cascades. Two different distributions are found: power-law and bimodal. Both distributions satisfy the condition of infrequent large events; nevertheless, they appear to be very different under many other aspects. Author found that the density of the network of interpersonal influences could be one explanation for these differences: if the network is sufficiently sparse the propagation of cascades is limited by the global connectivity of the network and cascades size exhibits a power-law distribution; on the contrary, when the network is sufficiently dense the propagation is limited by the stability of individual nodes and a bimodal distribution is observed.

In two different studies (15; 16), authors found that a small but significant fraction of posts forms wide and deep cascades and that different cascades may evolve in different ways (16), while in (15) authors presented a novel approach that examines the evolution of cascades over time and that is not limited to the prediction of the final size (and shape) from the initial conditions.

A fundamental aspect, yet to be explored, is that concerning the drivers of contents diffusion, both in the case of big cascades and single pieces of information. We will provide results on this problem in Chapter 3.

Misinformation

The **misinformation effect** was first exposed in an experimental work (55) that used neuroimaging to reveal its underling mechanisms. Memory

[1] A cascade is a chain of repeated action of the same kind – e.g., the chain of re-shares of a post.

distortions can result in pieces of misinformation, even without explicit external influence (56; 57). However, emotional states and beliefs affect human interpretation of the facts, often leading to the misinformation effect (58).

In the last years, the spreading of unsubstantiated and false claims through OSN (such as Facebook), that often reverberate in the online community, led to mass **digital misinformation**. Recent studies focused on how to stop misinformation diffusion by means of fact checking (59; 60) or algorithmic driven solutions (61; 62; 63) aimed at fighting misinformation. However, empirical results pointed out the inefficacy of such approaches on online social networks (21; 64).

Authors in (20) presented a detailed analysis of the information consumption by Facebook users on different categories of pages: alternative information sources, political activism, and main stream media. Authors pointed out evidences that main stream media information reverberate as long as unsubstantiated one, and that the exposition to the latter makes users more likely to interact with intentionally injected false information.

A similar study (65), based on scientific and conspiracy-like pages, reported the attitude of the users to split in separate communities each referring to one of the two categories of pages. While the consumption patterns for the information are similar, the conspiracy-like users are more prone to interact with posts and pages of their category and are especially prone to share the information, while scientific users tend to interact also with conspiracy posts. Moreover it is confirmed that conspiracy users are the ones more subject to interact with unverified information.

Chapter 3

The Spreading of Misinformation Online

A Data-Driven Model of Echo Chambers Formation

All the results shown in this chapter refer to the article (29)[1]. Our main interest lies in understanding the role of confirmation bias and homophily in shaping information cascades, and we investigate this aspect by means of a data-driven analysis of contents diffusion.

We address the information diffusion on Facebook focusing on two sets of distinct contents: (i) conspiracy theories and (ii) scientific information. Conspiracy theories simplify causation, reduce the complexity of reality, and contain uncertainty; scientific information disseminates scientific advances and exhibits the process of scientific thinking. The main difference between the two is content verifiability, while for scientific in-

[1]The results shown in this Chapter are all part of the paper (29), published in the Proceedings of the National Academy of Sciences. It is a joint work with Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Dr. Antonio Scala, Prof. Guido Caldarelli, Prof. H. Eugene Stanley, and Dr. Walter Quattrociocchi. MDV, WQ, and AS outlined the research question. MDV performed the simulations. FP downloaded the data. MDV, WQ, and AS interpreted the results. MDV, AB, FZ, FP, AS, GC, HES, and WQ contributed equally to the writing and reviewing of the manuscript.

formation a reference to peer-reviewed articles or institutional sources is usually available, the origins of conspiracy theories are often unknown and the content of the theories is strongly disengaged from mainstream society. The concept of content verifiability, and the attitude of users towards not verified information, is a topic of primary importance due to the massive diffusion of digital misinformation in online social media. As a matter of fact, the World Economic Forum (WEF) listed online misinformation as one of the main threats to our society (27; 28).

We analyze the cascades of information from the two separate samples and we find that homogeneity is the primary driver of the content diffusion and generates the formation of homogeneous, polarized clusters, i.e., *echo chambers* (20; 65; 66; 67), each of which has its own cascade dynamics. We also find that although consumers of scientific information and conspiracy theories exhibit similar consumption patterns with respect to the content, the cascade patterns of the two differ. Then we reproduce the dynamics of real data cascades through a percolation model of rumor spreading accounting for homogeneity and polarization.

3.1 Materials and Methods

3.1.1 Data Description

Using the approach described in Ref. (65), we define the space of our investigation with the support of diverse Facebook groups that are active in the debunking of misinformation.[2] We identify three main pages' categories according to the topics they promote: *conspiracy-like theories*, *science news*, and *trolling messages*. Conspiracy-like pages promote contents neglected by main stream media, science news pages diffuse scientific news and research advances for which it is easy to check the sources, while trolling messages pages include those pages that intentionally disseminate sarcastic false information on the Web. We download all the posts (and their respective user interactions) across a five-year timespan

[2]Protesi di Complotto, Che vuol dire reale, La menzogna diventa verità e passa alla storia.

(2010 to 2014) from a set of 69 public Facebook Italian pages divided according to our pages classification. In particular, information from troll pages are used as a benchmark to fit our data-driven model. We perform the data collection process by using the Facebook Graph API (68), which is publicly available and accessible through any personal Facebook user account. We get a total of $9,642$ news divided across the three information categories that can then be considered as sharing trees. For each news we download the successive sharing users' ids and sharing times by using a crawler for Facebook post trees through web browser automation (85). The crawler is able to download the complete sharing tree for each post in input and returns a file storing the edge list of the sharing tree and the respective sharing times. The exact breakdown of the data is presented in Tab. 1.

	Total	*Science*	*Conspiracy*	*Troll*
Pages	69	35	32	2
News	$9,642$	$5,032$	$3,538$	$1,072$
Labeled Users	$73,379$	$14,613$	$58,766$	–
Shares	$266,211$	$59,059$	$181,914$	$25,238$

Table 1: Data Description.

3.1.2 Basic Definitions

The best way to represent our news data is by considering them as *trees of information sharers*, that are otherwise connected by an underlying network structure. In our particular case, the underlying network structure is the Italian subset of the Facebook friendship network that we represent as a scale-free network. We introduce some notations that we will extensively use throughout the chapter.

Sharing Tree. A tree is an undirected simple graph that is connected and has no simple cycles. An oriented tree is a directed acyclic graph whose underlying undirected graph is a tree. A sharing tree in the context of our research is an oriented tree made up of the successive sharing

occurrences of a news item through the Facebook system. The root of the sharing tree is the node that performs the first temporal share. We define the *size* of the sharing tree as the number of nodes (and hence the number of news sharers) in the tree and the *height* of the sharing tree as the maximum path length starting from the root.

User Polarization. We define the user polarization $\sigma_i = 2\varrho_i - 1$, where $0 \leq \varrho_i \leq 1$ is the fraction of likes a user i made on conspiracy related content and $i \in \{1, \ldots, N\}$. Notice that $-1 \leq \sigma_i \leq 1$. User i is said to be polarized in science (resp. conspiracy) if $\sigma_i \leq -0.95$ (reps. if $\sigma_i \geq 0.95$).

Edge Homogeneity. From the user polarization σ_i, we define the edge homogeneity for any edge e_{ij} between nodes i and j, as

$$\sigma_{ij} = \sigma_i \sigma_j,$$

with $-1 \leq \sigma_{ij} \leq 1$. Edge homogeneity reflects the similarity level between the polarization of the two sharing nodes. A link in the sharing tree is homogeneous if its edge homogeneity is positive, otherwise it is non homogeneous. We then define a sharing path to be any path from the root to one of the leaves of the sharing tree. A *homogeneous path* is a sharing path for which the edge homogeneity of each edge is positive, i.e., a sharing path whose edges are all homogeneous links.

Wald Test. We use the Wald test to compare the scaling parameters of two power law distributions. We define it as

$$H_0 : \hat{\alpha_1} = \hat{\alpha_2}$$
$$H_1 : \hat{\alpha_1} \neq \hat{\alpha_2}$$

where $\hat{\alpha_1}$ and $\hat{\alpha_2}$ are the estimated scaling parameters. The Wald statistic:

$$W = \frac{(\hat{\alpha_1} - \hat{\alpha_2})^2}{Var(\hat{\alpha_1})},$$

follows a χ^2 distribution with one degree of freedom. We reject the null hypothesis H_0 and conclude that there is a significant difference between

the two scaling parameters if the p-value of W is below a given significance level α.

Kolmogorov-Smirnov Test. We use the Kolmogorov-Smirnov test to compare the empirical distribution functions of two samples.

The Kolmogorov-Smirnov statistic for two given cumulative distribution functions $F_1(x)$ and $F_2(x)$ is

$$D = \sup_x |F_1(x) - F_2(x)|,$$

which measures the maximum punctual distance between the two sample distributions. If D is bigger than a given critical value D_α[3] we reject the null hypothesis $H_0 : F_1(x) = F_2(x)$ and conclude that there is a significant difference between the two sample distributions.

3.2 Analysis of Observable Data

In this section we analyze the sharing trees data in order to characterize the news sharing patterns of Facebook. We first study the basic properties (size, height, max degree, and mean degree) of sharing trees for the three categories –i.e., science news, conspiracy theories, and trolling messages. In Fig. 1 we report the complementary cumulative distribution function (CCDF) of size (Fig. 1(a)), the cumulative distribution function (CDF) of height (Fig. 1(b)), the CCDF of maximum degree (Fig. 1(c)), and the CCDF of mean degree (Fig. 1(d)) for all categories. The distribution of the size of the sharing trees and that of their maximum degree are power low. The estimated exponents for the distribution of size are 2.21, 2.47, 2.44 and those for the distribution of the maximum degree

[3]The critical value D_α depends on the sample sizes and on the considered significance level α, it can be computed as

$$D_\alpha = c(\alpha)\sqrt{\frac{n_1 + n_2}{n_1 n_2}},$$

where n_1 and n_2 are the respective sample sizes and $c(\alpha)$ is a fixed value associated with the significance level α.

are 2.16, 2.45, 2.41, respectively for science news, conspiracy-like theories, and trolling messages. We note a similar behavior, both for size and maximum degree, for all categories. From Fig. 1(c), we can see that the maximum height reached is 5 for science news and conspiracy-like theories, and 4 for trolling messages, while for all categories there is a high probability that the height of the sharing tree remains below 3. The mean degree instead is with high probability smaller than 10, for all categories.

Cascade lifetime is defined as the number of hours between the first user and the last user sharing a post. We compare the lifetime of science and conspiracy information cascades and we report their probability density functions (PDFs) in Fig. 2. For both categories we find a first, high, peak of probability in the first two hours after the first sharing of the post and a second one after ~ 20 hours. We deduce that the temporal sharing patterns are similar irrespective of the difference in topic. There is an high probability that a post won't last more that few hours, indeed, we find that a significant percentage of the information diffuses rapidly (24.42% of the science news and 20.76% of the conspiracy-like theories diffuse in less than two hours, and 39.45% of science news and 40.78% of conspiracy-like theories in less than five hours). Only 26.82% of the diffusion of science news and 17.79% of conspiracy-like theories lasts more than one day. Kolmogorov-Smirnov test makes us reject the hypothesis H_0 that the two distributions are equal.

One measure that might influence the cascade's lifetime is its size. For both science news and conspiracy-like theories, we show the cascade's lifetime as a function of its size in Fig. 3(a). We note that the news assimilation is content-driven. In the specific, for science news we have a peak in the lifetime corresponding to a cascade size value of ≈ 200, and higher cascade size values correspond to high lifetime variability. For conspiracy related content the lifetime increases with cascade size. While for science news a longer lifetime does not correspond to a higher level of interest, conspiracy rumors show a positive relation between lifetime and size. As a control we also compute the cascade's size as a function of its lifetime (Fig. 3(b)). We confirm the existence of a differentiation in the sharing patterns that is content-driven. We also confirm that

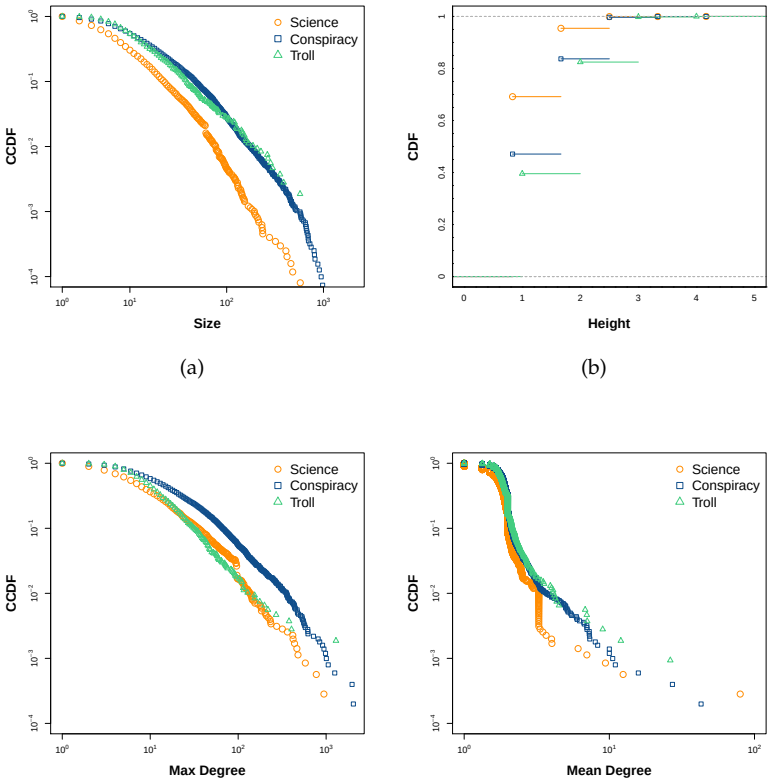


Figure 1: CCDF of Size (a), CDF of Height (b), CCDF of Maximum Degree (c), and CCDF of Mean Degree (d). Size and max degree show power law distributions, where the estimated exponents for the power law distribution of size are 2.21, 2.47, 2.44 and those of max degree are 2.16, 2.45, 2.41, respectively for science news, conspiracy-like theories, and trolling messages. Height is generally low, with the maximum level being 5 for science news and conspiracy-like theories, and 4 for trolling messages.

for conspiracy-like theories there is a positive relation between cascade's size and lifetime.

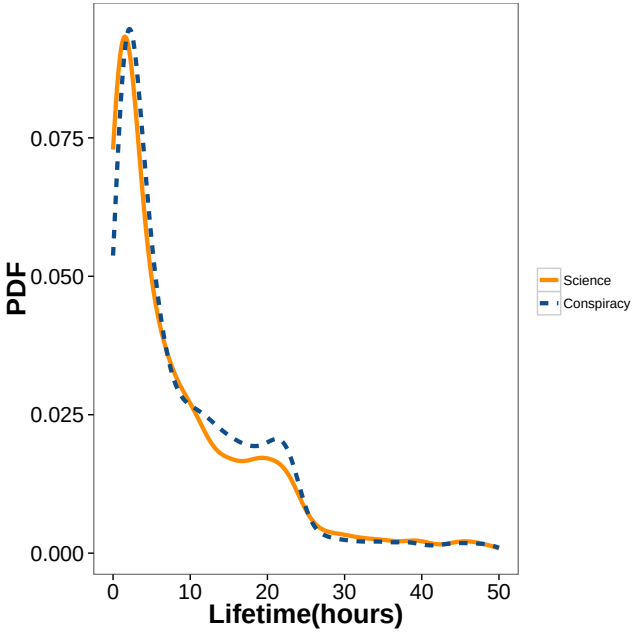
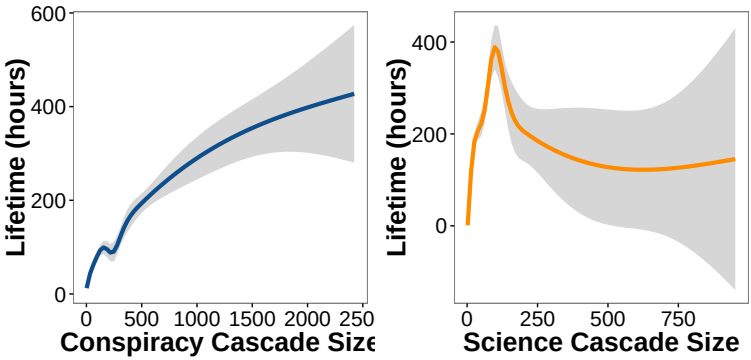
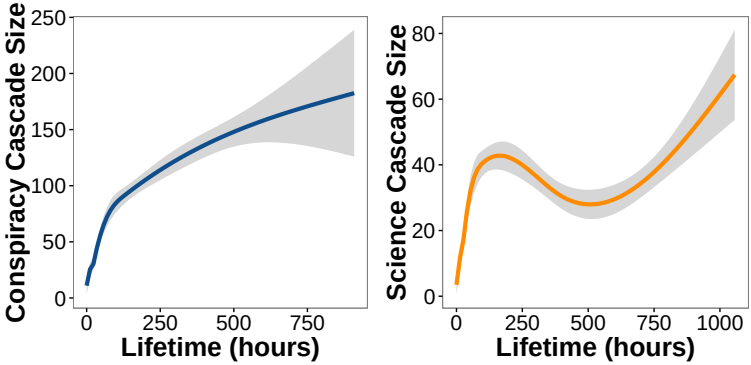


Figure 2: Probability density functions (PDFs) of cascades lifetime for science (solid orange) and conspiracy (dashed blue). The lifetime is here computed as the difference of hours between the first and last share of a post. Both categories show a similar behavior, with a peak in the first two hours and another one after about 20 hours.



(a)



(b)

Figure 3: (a) Lifetime as a function of the cascade size for conspiracy-like theories (left) and science news (right). We note a content-driven differentiation in the sharing patterns. For conspiracy-like theories the lifetime grows with the size, while for science news there is a peak in the lifetime around a value of the size equal to 200, and a higher variability in the lifetime for larger cascades. (b) Cascade size as a function of the lifetime for conspiracy-like theories (left) and science news (right).

The main focus of this study is to characterize the sharing patterns of two specific information categories on Facebook. Identifying the drivers of information cascades and understanding the mechanism of news resharing is especially necessary when dealing with false or not verified information, that could potentially be very harmful. Figure 4 shows the PDF of the mean edge homogeneity, computed for all cascades of science news and conspiracy-like theories. We notice a prevalence of homogeneous links between consecutively sharing users. In particular, the average edge homogeneity value of the entire sharing cascade is always greater than or equal to zero, indicating that either the information transmission occurs inside homogeneous clusters in which all links are homogeneous or it occurs inside mixed neighborhoods in which the balance between homogeneous and non homogeneous links is favorable towards the former ones. However, the probability of close to zero mean edge homogeneity is really small.

To further characterize the role of homogeneity in shaping sharing cascades, we compute the cascade size as a function of the mean edge homogeneity for both science news and conspiracy-like theories, see Fig. 5. For science news, higher levels of mean edge homogeneity in the interval (0.5, 0.8) correspond to larger cascades, but for conspiracy-like theories lower levels of mean edge homogeneity (~ 0.25) correspond to larger cascades. Notice that, although viral patterns related to distinct contents differ, homogeneity is clearly the driver of information diffusion. In other words, different contents generate different echo chambers, characterized by the high level of homogeneity inside them.

To completely understand the structure of the sharing trees, for each tree we look at all its paths that start from the root. We recall that a homogeneous path is a path for which the edge homogeneity of all its edges is positive, meaning that every two consecutive sharing users are polarized towards the same category. Figure 6 shows the complementary cumulative distribution function (CCDF) of the number of homogeneous and total paths, for the three following samples: science news and conspiracy-like theories together (left), science news (center), and conspiracy-like theories (right). More formally, we consider, for each tree, the number of

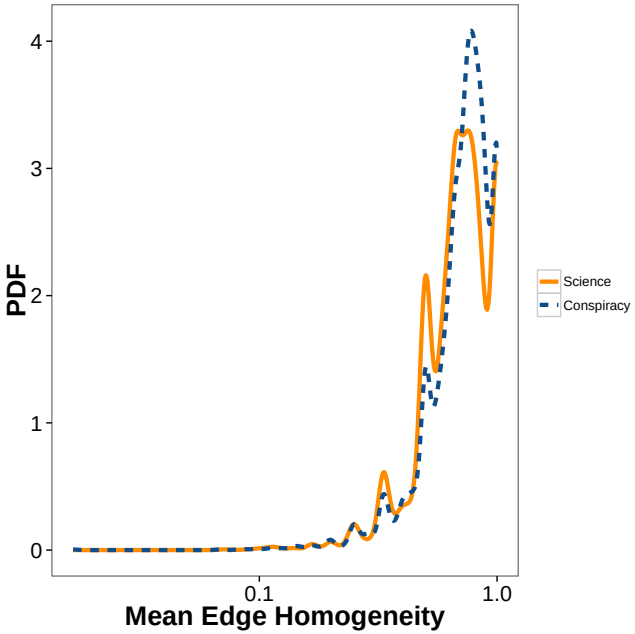


Figure 4: Mean edge homogeneity for science news (solid orange) and conspiracy-like theories (dashed blue). The mean value of edge homogeneity on the whole sharing cascades is always greater than or equal to zero.

all sharing paths from the root to one of the leaves, and we compare it with the number of sharing paths with positive edge homogeneity. Looking at Fig. 6 we notice a high similarity for all the couples, for this reason we compare them by using Wald test and Kolmogorov-Smirnov test, with level of significance $\alpha = 0.01$. The null hypothesis is the equivalence of the two scaling parameters for the Wald test and the equivalence of the whole sample distributions for the Kolmogorov-Smirnov. Table 2 reports the results from Wald test, while Table 3 those from Kolmogorov-Smirnov test.

We fail to reject the null hypothesis of Wald test in the second and third case, i.e., science news and conspiracy theories separately, while we

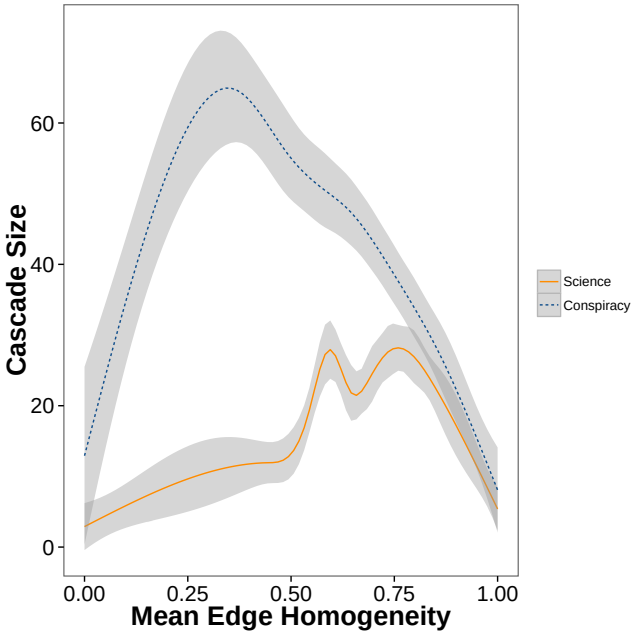


Figure 5: Cascade size as a function of mean edge homogeneity for science news (solid orange) and conspiracy-like theories (dashed blue).

reject it in the case of the whole sample. However, we fail to reject the null hypothesis of Kolmogorov-Smirnov in all three cases, as the maximum distance is always smaller than the critical value, we deduce that the distributions are not significantly statistically different in all three cases, and the same is true for the scaling parameters in the case of science news and conspiracy theories separately. The similarity between each pair of distributions is well captured by the Q-Q plots shown in Fig. 7.

The temporal sharing patterns of science news and conspiracy-like theories are similar irrespective of the contents. Indeed, for both categories, cascade's lifetime exhibits a first peak of probability in the first two hours, a second one around a lifetime of 20 hours, and after that the probability rapidly decreases. Despite the similar consumption patterns,

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	p
I	2.037	2.089	8.45×10^{-7}
II	2.427	2.447	0.413
III	2.054	2.026	0.040

Table 2: Results from Wald test, where $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are the two estimated scaling parameters for each couple and p is the corresponding p-value.

	D	D_α	p
I	0.0216	0.0233	0.02047
II	0.0199	0.0378	0.4525
III	0.0262	0.0296	0.03204

Table 3: Results from Kolmogorov-Smirnov test. D is the estimated maximum distance between the two distributions under analysis, D_α is the corresponding critical value, and p the corresponding p-value.

cascade lifetime expressed as a function of cascade size differs greatly for the different content sets. The PDF of the mean edge homogeneity indicates that homogeneity is pervasive in the linking step of sharing cascades. Moreover, we observed that the distribution of the number of total and homogeneous sharing paths are very similar for both content categories. Viral patterns related to contents belonging to different narratives differ, but homogeneity is clearly the driver of content diffusion.

3.3 The Model

In the previous section we proved that homogeneity can be considered the primary driver of information spreading on Facebook, when the information relates to highly polarized communities. Accounting for both homogeneity and polarization, we now introduce a data-driven model of rumor spreading. Let n be the number of users that are connected by a small-world network with rewiring probability r (69), and consider a news set of size m. Every user has an opinion ω_i, $i \in \{1, n\}$, uniformly distributed in $[0, 1]$ and every news has a fitness (degree of in-

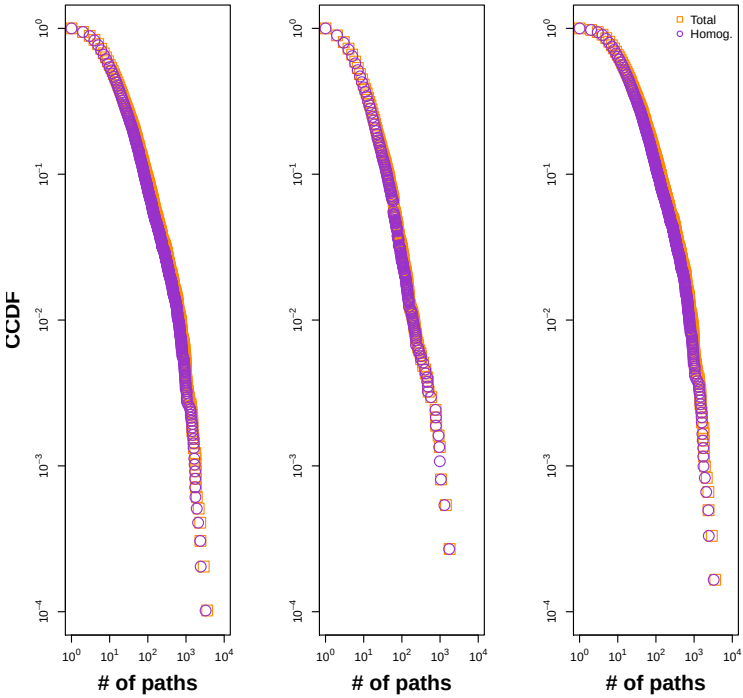
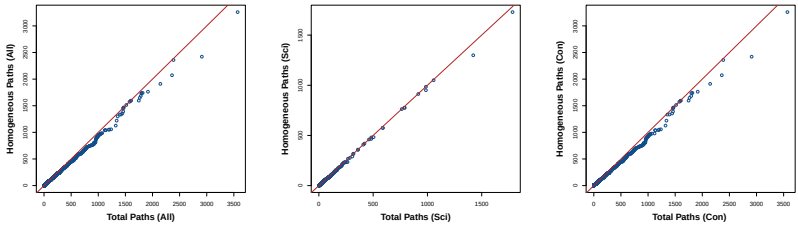


Figure 6: Complementary cumulative distribution function (CCDF) of the number of total and homogeneous paths on the whole science and conspiracy sample (left), on science news (center), and on conspiracy-like theories (right). Results from the Kolmogorov-Smirnov test on the three distributions couples, with the null hypothesis H_0 that the distributions in each couple are equal, show the following p-values: 0.02, 0.45, 0.03, leading us to reject the null hypothesis in all cases ($\alpha = 0.01$). On the other hand the maximum estimated distances are respectively 0.0216, 0.0199, 0.0262 and corresponding critical values 0.0232, 0.0377, 0.0296; the maximum distance is smaller than the critical value in all cases, so we fail to reject the null hypothesis. We can consider the distributions in each couple as equal, meaning that homogeneous paths are pervasive.

terest) ϑ_j, $j \in \{1, m\}$, uniformly distributed in $[0, 1]$. In order to mimic the diffusion of the news among the population, at each step one of the



(a) (b) (c)

Figure 7: Q-Q plots for the three couples of distributions in Fig. 6: number of total vs homogeneous links on science news and conspiracy theories together (a), on science news (b), and on conspiracy-like theories (c). In all three cases the distributions can be considered as equal.

news items are initially shared by a group of first sharers. After the first share, the news recursively passes to the neighbors of previous step sharers, that in turn share the news only if their opinion is close enough to the fitness of the news, i.e., user i shares news j if and only if:

$$|\omega_i - \vartheta_j| \leq \delta,$$

where δ is the sharing threshold.

We assume that the flow of information only passes through like minded users; to reflect this property, also observed in the real data sample, we model the connectivity pattern as a signed network. We define ϕ_{HL} as the fraction of homogeneous links in the network, let M be the number of total links and n_h the number of homogeneous ones, then we have:

$$\phi_{HL} = \frac{n_h}{M},\ 0 \leq n_h \leq M.$$

Notice that $0 \leq \phi_{HL} \leq 1$ and that $1 - \phi_{HL}$, the fraction of non homogeneous links, is complementary to ϕ_{HL}. In particular, we can reduce the parameters space to $\phi_{HL} \in [0.5, 1]$ without loss of generality, as we would restrict our attention to either one of the two complementary clusters.

The model can be seen as a branching process where the sharing threshold δ and neighborhood dimension z are the key parameters. More formally, let the fitness θ_j of the j^{th} news and the opinion ω_i of a the i^{th} user be uniformly i.i.d. in $[0, 1]$. Then the probability that a user i shares a post j is defined by a probability $p = \min(1, \theta + \delta) - \max(0, \theta - \delta) \approx 2\delta$, since θ and ω are uniformly i.i.d. In general, if ω and θ have distributions $f(\omega)$ and $f(\theta)$, then p will depend on θ,

$$p_\theta = f(\theta) \int_{\max(0,\theta-\delta)}^{\min(1,\theta+\delta)} f(\omega)\, d\omega.$$

If we assume that the distribution of the number m of the first sharers is $f(m)$, then the average cascade size is

$$S = \sum_m f(m)\, m\, (1-\mu)^{-1} = \frac{\langle m \rangle_f}{1-\mu} \approx \frac{\langle m \rangle_f}{1-2\delta z},$$

where $\langle \ldots \rangle_f = \sum_m \ldots f(m)$ is the average with respect to f. In the simulations we fixed a neighborhood dimension $z = 8$ as the branching ratio μ depends on the product of z and δ and, without loss of generality, we can consider the variation of just one of them.

3.4 Simulation Results

A preliminary challenge arises in the first step of the model, when we need to specify the first sharers distribution. The natural choice would be the data distribution, and we use it as a benchmark to compare our choices and as a sample to fit the parameters. In turn we compare our sample distribution with: (i) the Inverse Gaussian (IG), (ii) the Log-normal (LN), (iii) the Poisson (Poi), (iv) and the Uniform (U) distribution. All parameters are chosen to fit the real data distribution from both the science and conspiracy news sample, and the trolling messages sample. In Table 4 we show a summary of relevant statistics (min value, first quantile, median, mean, third quantile, and max value) for the fitted distributions and we compare them to the real data first sharers distribution, for both the science and conspiracy news sample, and the trolling

messages sample, while in Figure 8 we show the PDF of the best fitting function for the case of science news and conspiracy-like theories. The inverse Gaussian (IG) with mean 39.34 and scale parameter 6.28 shows the best fit with respect to all the considered statistics for the distribution of first shares of science news and conspiracy-like theories data, and again the IG, with mean 18.73 and scale parameter 9.63, shows the best fit for the distribution of first sharers from trolling messages category.

We perform two sets of Monte Carlo simulations of the model that differentiate themselves in the amount of real data information they exploit. The first one only takes as data the fitted parameters for the first shares distribution (with respect to the science news and conspiracy-like theories sample), while the second one also assume a number of agents and news items equal to those of the trolling messages dataset. Along with the first sharers distribution, we vary the sharing threshold δ in the interval $[0.01, 0.05]$ and the fraction of homogeneous links ϕ_{HL} in the interval $[0.5, 1]$. To avoid biases induced by statistical fluctuations in the stochastic process, each point of the parameter space is averaged over 100 iterations. A good estimate of real data values is provided by an almost even fraction of homogeneous links, $\phi_{HL} \sim 0.5$. This result points out that the network is divided in two clusters inside which the news items remain isolated and can only be transmitted within each community's echo chamber.

Science and Conspiracy

	Data	*IG*	*LN*	*Poi*
Min	*1*	**0.36**	0.10	20
1st Qu.	*5*	**4.16**	3.16	35
Median	*10*	**10.45**	6.99	39
Mean	*39.34*	**39.28**	13.04	39.24
3rd Qu.	*27*	**31.59**	14.85	43
Max	*3033*	**1814**	486.10	66

Troll

	Data	*IG*	*LN*	*Poi*
Min	*1*	**0.81**	0.16	*7*
1st Qu.	*5*	**4.67**	2.13	16
Median	*8*	**9.71**	4.66	19
Mean	*18.73*	**18.73**	9.99	18.72
3rd Qu.	*16*	**21.76**	11.85	22
Max	*3882*	**346.60**	183.40	32

Table 4: On the top part of the table we report a summary of relevant statistics (minimum value, first quantile, median, mean, third quantile, and maximum value) for real data, from science news and conspiracy-like theories, and fitted ($IG(39.33, 6.27)$, $LN(2.47, 1.40)$, $Poi(39.34)$) first sharers' distributions. On the bottom part we report the same statistics for real data, from trolling messages, and fitted ($IG(18.73, 9.63)$, $LN(2.21, 0.93)$, $Poi(18.73)$) first sharers' distributions.

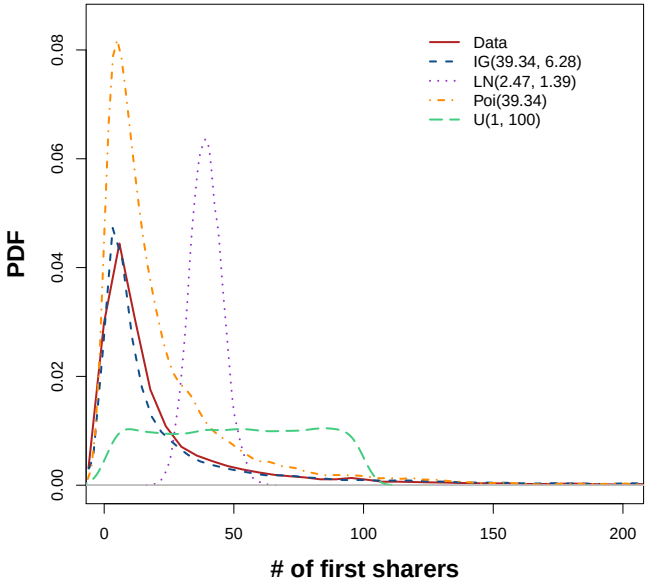


Figure 8: Probability density function (PDF) of real data from the science and conspiracy news sample (solid red) and fitted first sharers distributions: Inverse Gaussian (dashed blue), Log Normal (dotted violet), Poisson (dot-dashed orange), and Uniform in $[0, 1]$ (dashed green).

Figure 9 shows the results obtained by the first set of simulations, for $n = 5,000$ users, $m = 1,000$ news, a varying fraction of homogeneous links $\phi_{HL} \in [0.5, 1]$, the rewiring probability $r \in [0.01, 0.2]$, and the number of first sharers distributed in turn as $IG(39.33, 6.27)$, $LN(2.46892, 1.39399)$, $Poi(39.3385)$, and $U(0, 1)$. Figure 10 shows the results for the second set of simulations, relative to the fit of the model on the trolling messages dataset, where the number of users and the number of messages is taken from the trolling set data ($n = 16,889$ is the number of users active in the trolling category and $m = 1,072$ is the number of trolling messages in the dataset), the parameters ϕ_{HL} and r vary in the same intervals as before, and the number of first sharers is in turn distributed as $IG(18.73, 9.63)$, $LN(2.21, 0.93)$, and $Poi(18.73)$. For both figures the different colors and shapes of the points indicate the different distributions of first sharers used in the simulations: red square is for the sample data distribution, blue circle for the Inverse Gaussian distribution, violet triangle for the Log-Normal distribution, orange cross for the Poisson distribution, and green x for the Uniform distribution. Comparing the results obtained with synthetic first sharers distribution with those obtained using the real data one, we see that the IG distribution provides the best fit for all parameters choices.

In order to validate our model, we also vary the parameter $\delta \in [0.01, 0.05]$ for the second set of simulations, with first sharers distribution fitted on trolling messages data, $n = 16,889$ users, and $m = 1,072$. Fig. 11 shows the simulated average size (left) and average height (right) of the sharing trees, where different colors and shapes of the points indicate the different distributions of the number of first sharers considered, and we vary the sharing threshold δ and the fraction of homogeneous links ϕ_{HL}, i.e., $\delta \in [0.01, 0.05]$ and $\phi_{HL} \in [0.5, 0.59]$. In Table 5 we report the combination of parameters that best reproduces the real data values and the mean and standard deviation of cascades' size and height for that combination of parameters. The assumptions for the simulations are the same as in Fig. 11. We compare the real data from trolling messages dataset with simulated values obtained using IG or real data as first sharers distribution.

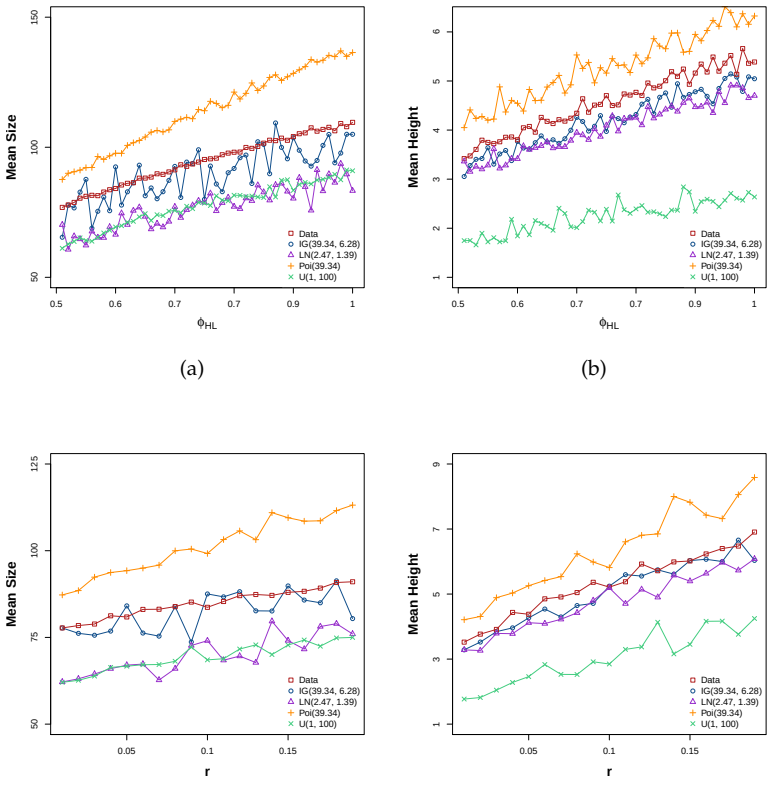


Figure 9: Simulation results for the mean size (a,c) and the mean height (b,d) of news sharing trees, where the network contains $n = 5,000$ users and $m = 1,000$ news are shared in turn. We fix the sharing threshold $\delta = 0.05$ and test different combinations of fraction of homogeneous links ϕ_{HL} and rewiring probability r. In Fig. 9(a-b) $r = 0.01$ is fixed and ϕ_{HL} varies in the interval $[0.5, 1]$; while in Fig. 9(c-d) $\phi_{HL} = 0.5$ is fixed and r varies in the interval $[0.01, 0.2]$. The different colors and shapes of the points indicate the different distributions of first sharers used in the simulations: red square is for the sample data distribution, blue circle for the $IG(39.33, 6.27)$, violet triangle for the $LN(2.46892, 1.39399)$, orange cross for the $Poi(39.3385)$, and green x for the $U(0, 1)$.

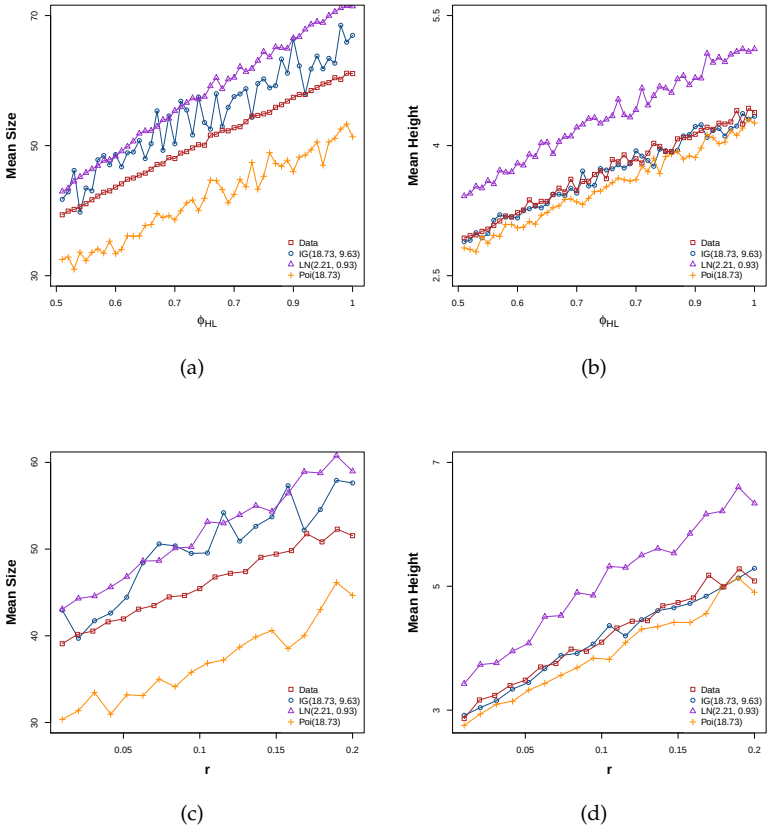


Figure 10: Fit of the model with trolling messages data, simulation results for the mean size (a,c) and the mean height (b,d) of news sharing tree, where the networks contains $n = 16,889$ users and $m = 1,072$ news are shared. The values of n and m are equal to the real data ones for the trolling messages data. We fix the sharing threshold $\delta = 0.05$ and test different combinations of fraction of homogeneous links ϕ_{HL} and rewiring probability r. In Fig. 10(a-b) $r = 0.01$ is fixed and ϕ_{HL} varies in the interval $[0.5, 1]$; while in Fig. 10 (c-d) $\phi_{HL} = 0.5$ is fixed and r varies in the interval $[0.01, 0.2]$. The different colors and shapes of the points indicate the different distributions of first sharers used in the simulations: red square is for the sample data distribution, blue circle for the $IG(18.73, 9.63)$, violet triangle for the $LN(2.21, 0.93)$, and orange cross for the $Poi(18.73)$.

	Param.	*Size*		*Height*	
	(ϕ_{HL}, δ)	*Mean*	*Std Dev.*	*Mean*	*Std Dev.*
RD	-	23.54	122.32	1.78	0.73
S-RD	.56, .015	23.52	133.02	1.28	1.18
S-IG	.56, .015	23.42	33.43	1.28	0.88

Table 5: Mean cascades size and height obtained with the best parameter combination compared to real data measures (RD). Simulation results are reported for two cases: number of first sharers distributed as real data (S-RD) and as inverse Gaussian (S-IG).

Taking into account the results in Fig. 11 and in Tab. 4-5, we perform a last simulation of the model dynamics with the first sharers distributed as $IG(18.73, 9.63)$, $n = 16,889$, $m = 1,072$, and the best observed combination of the other model parameters $(\phi_{HL}, r, \delta) = (0.56, 0.01, 0.015)$. The CDF of sharing trees height and the CCDF of size is reported in Fig. 12. Both measures show a good fit with data. Table 6 shows a summary of relevant statistics (min value, first quantile, median, mean, third quantile, and max value) to compare the real data size and height distributions with the fitted ones, for the same distribution of first sharers and the same parameters combination. We notice that the fit is good for all the statistics, with the exception of min and max value of size. For the min value, the presence of a zero is due to the fact that the Inverse Gaussian is a real valued distribution function and in the simulations we considered the integer part of the number of first sharers, thus producing a number of never shared pieces of information. On the other hand, the high difference in the max value is probably due to the long tail of the data size distribution.

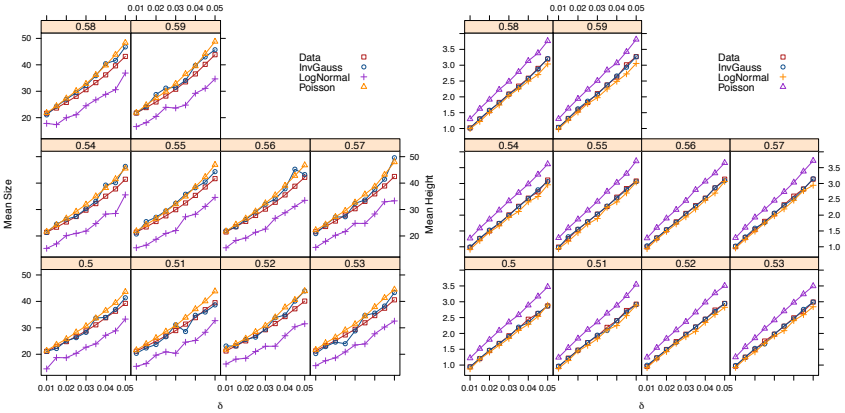


Figure 11: Fit of the model with trolling messages data, simulation results for the mean size (left) and the mean height (right) of the sharing trees, with $n = 16,889$ users, $m = 1,072$ news, fixed rewiring probability $r = 0.01$, fraction of homogeneous links ϕ_{HL} varying in the interval $[0.5, 0.59]$, and sharing threshold δ varying in $[0.01, 0.05]$. The different colors and shapes of the points indicate the different distributions of first sharers used in the simulations.

	Size		***Height***	
	Data	*Simulated*	*Data*	*Simulated*
Min	2	0	1	1
1st Qu.	7	6.19	1	2
Median	10	11.93	2	2
Mean	23.54	23.84	1.78	2.18
3rd Qu.	19.25	27.17	2	3
Max	3845	541.80	4	5

Table 6: Summary of relevant statistics for size and height distributions.

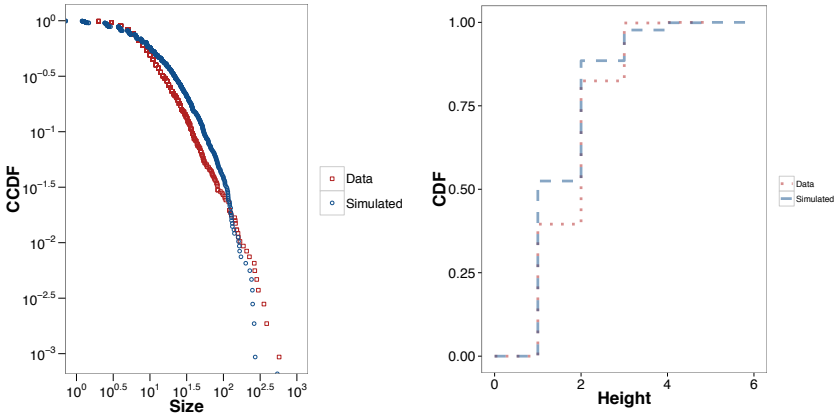


Figure 12: Complementary cumulative distribution function (CCDF) of size (left) and cumulative distribution function (CDF) of height (right) for the best parameters combination that fits troll data values, $(\phi_{HL}, r, \delta) = (0.56, 0.01, 0.015)$, and first sharers distributed as $IG(18.73, 9.63)$. We note that it is indeed a good fit of trolling data.

3.5 Concluding Remarks

Whether a news item, either substantiated or not, is accepted as true by a user may be strongly affected by social norms or by how much it coheres with the user's system of beliefs (58; 70). Despite enthusiastic claims that social media is generating a vast "collective intelligence" available to all (71), many mechanisms cause false information to gain acceptance, which in turn generate false beliefs that, once adopted by an individual, are highly resistant to correction (72; 73; 74; 75). Using extensive quantitative analysis we show that social homogeneity is the primary driver of content diffusion, and one frequent result is the formation of homogeneous, polarized clusters (often called "echo chambers"). We also find that although consumers of science news and conspiracy-like theories show similar consumption patterns with respect to content, their cascades differ. Social homogeneity appears to be the primary driver of content diffusion, and each echo chamber has its own cascade dynamics. To mimic these dynamics, we introduce a data-driven percolation model of signed networks, i.e., networks composed of signed edges. Our analysis shows that for science news and conspiracy-like theories the cascade lifetime has a probability peak in the first two hours followed by a rapid decrease. Although the consumption patterns are similar, cascade lifetime as a function of the size differs greatly. The PDF of the mean edge homogeneity indicates that homogeneity is present in the linking step of sharing cascades. The distribution of the number of total sharing paths and homogeneous sharing paths are similar in both content categories. Viral patterns related to distinct contents are different but homogeneity drives content diffusion. We simulate our data-driven percolation model by fixing the number of users and news items downloaded from troll pages and varying the other parameters. We compare the simulated results with the data and find a high level of similarity.

Chapter 4

Echo Chambers: Emotional Contagion and Group Polarization on Facebook

All the results shown in this chapter refer to the submitted paper (30)[1]. We address the relationship between group polarization and confirmation bias by quantitatively analyzing the temporal evolution of two online communities in the Italian Facebook, linked to either scientific or conspiracy-like contents. Moreover we link the group polarization to the emotional contagion inside the echo chambers by means of statistical and sentiment analysis techniques.

Misinformation has traditionally represented a political, social, and economic risk. The digital age, in which new ways of communication arose, has exacerbated its extent, and mitigation strategies are even more uncertain. However, according to the World Economic Forum, massive digital misinformation remains one of the main threats to our so-

[1]The results shown in this Chapter are all part of the paper (30), available as a pre-print at arXiv:1607.01032. It is a joint work with Dr. Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Dr. Antonio Scala, Prof. Guido Caldarelli, and Dr. Walter Quattrociocchi. MDV, GV and WQ outlined the research question. MDV and GV provided the analysis tools. MDV, GV, and WQ performed the analysis and interpreted the results. MDV, GV, AB, FZ, AS, GC, and WQ contributed equally to the writing and reviewing of the manuscript.

ciety (28).

The diffusion of social media caused a shift of paradigm in the creation and consumption of information. We passed from a mediated (e.g., by journalists) to a more disintermediated selection process. Such a disintermediation elicits the tendencies of the users to a) select information adhering to their system of beliefs – i.e., confirmation bias – and b) to form groups of like-minded people where they polarize their opinion – i.e. echo chamber (76; 77; 78; 79; 80; 81).

Under these settings, discussion within like-minded people seems to negatively influence users' emotions and to enforce group polarization (82; 83). What's more, experimental evidence shows that confirmatory information gets accepted even if containing deliberately false claims (22; 29; 65; 84; 85), while dissenting information is mainly ignored or might even increase group polarization (64). Current solutions, such as debunking efforts or algorithmic driven solutions based on the reputation of the source, seem to be ineffective. To make things more complicated, users on social media aim at maximizing the number of likes and often information, concepts, and debate get flattened and oversimplified.

In such a disintermediated environment, indeed, the public opinion deals with a large amount of misleading information generated by nationalists, populists, and conspirators, that is corrupting reliable sources at the heart. Computational social science (1) seems to be a powerful tool for a better understanding of the cognitive and social dynamics behind misinformation spreading (28). Along this path, in the (30) we address the evolution of online echo chambers by performing a comparative analysis of two distinct polarized communities on the Italian Facebook, i.e., science and conspiracy.

This work aims to study, quantitatively characterize, and model both the process of spreading news and the consumption of news for the early detection of trends in public opinion. The sizes of both the communities are firstly analyzed in terms of their temporal evolution and fitted by classical population growth models deriving from biology and medicine fields. The behavior of users turns out to be similar for both categories, irrespective of the contents: both science and conspiracy communities

reach a thresholding value in their sizes, after an almost exponential growth, in agreement with classical growth models.

Moreover, we analyze the community behavior by accounting for the engagement and the emotional dynamics of users. Indeed, whether a news item, either substantiated or not, is accepted as true by a user may be strongly affected by social norms or by how much it coheres with the community shared system of beliefs.

Users' emotional behavior seems to be affected by their engagement within the community. An higher involvement in the echo chamber, resolves in a more negative emotional state. Such a phenomenon appears in both users categories. Moreover, we observe that, on average, more active users show a faster shift towards the negativity than less active ones.

The chapter is structured as it follows. In section 4.1 we describe the data collection and give an account of the methodologies used. In section 4.2 we analyze the structural evolution of both science and conspiracy communities on the Italian Facebook. Then, in section 4.3 we explore the sentiment behavior of users as single units, and subsequently we explore the sentiment contagion inside each of the two communities from a macroscopic point of view.

4.1 Materials and Methods

Data Collection and Description

Using the approach described in (65), with the support of diverse Facebook groups very active in the debunking of misinformation (Protesi di Complotto, Che vuol dire reale, La menzogna diventa verità e passa alla storia), we identified two main categories of pages: *conspiracy-like theories*, i.e., pages promoting contents neglected by main stream media, and *science information*, i.e., pages diffusing scientific news and research advances for which it is easy to check the sources. Starting from this basic differentiation, we categorized Facebook pages according to their contents and their self description. The resulting dataset is composed of 73

public Italian Facebook pages, 34 of which were diffusing scientific information and 39 conspiracy-like theories, and covers a timespan of 5 years, from 2010 to 2014. Table 7 summarizes the details of our data collection.

	Total	*Science*	*Conspiracy*
Pages	73	34	39
Posts	271,296	62,705	208,591
Likes	9,164,781	2,505,399	6,659,382
Comments	1,017,509	180,918	836,591

Table 7: Dataset description.

The data collection process has been carried out using the Facebook Graph application program interface (API) (68), which is publicly available. For the analysis (according to the specification settings of the API) we only used publicly available data (thus users with privacy restrictions are not included in the dataset). The pages from which we download data are public Facebook entities and can be accessed by anyone. User content contributing to these pages is also public unless the users privacy settings specify otherwise, and in that case it is not available to us.

Growth Models

We fit the temporal evolution of the size of the two communities under analysis by means of the three classical growth models: the *Gompertz Growth* model, the *Logistic Growth* model, and the *Log-Logistic Growth* model.

The **Gompertz Growth Model** is often used to model growth phenomena which are typically characterized by an asymptotic behavior rather than by a linear increase. In that sense, a Gompertz function has to be intended as a special case of the most general logistic function, and it is nowadays applied in various research fields, such as biology, ecology, economics, marketing, and medicine. In oncology, in particular, the Gompertz sigmoid function has been used to model tumor growth (86; 87), which are interpreted as an expansion of cellular populations developing in a confined space, where the availability of nutrients

is limited in a certain sense. As a consequence, the model considers two parameters, a first one, a, for the tumor intrinsic growth related to the mitosis rate and a second one, b, for the growth deceleration, due to the antiangiogenic processes. Let $x(t)$ be the size of the tumor at time t, then we have:

$$\frac{dx(t)}{dt} = ax(t) - bx(t)\ln x(t).$$

For a given initial condition $x(0) = x_0$, and known parameters a and b, the solution (86) is:

$$x(t) = e^{a/b + (\ln(x_0) - a/b)e^{-bt}}. \tag{4.1}$$

The most general **Logistic Growth Model** is defined as:

$$x(t) = c + \frac{d-c}{(1+e^{b(t-f)})^g}. \tag{4.2}$$

In that case, the first stage of the growth is approximately exponential, then the growth rate decreases till an asymptotic value is reached. The right-hand asymptote is reached less gradually than the left-hand one compared to the behavior of the Gompertz function. We used two variants of the Logistic model to fit our data: $L3$ that considers only parameters (b, d, f) in (4.2), and $L5$ that is exactly (4.2).

Finally, the **Log-Logistic Growth Model** is defined as:

$$x(t) = c + \frac{d-c}{(1+e^{b(\ln t - \ln f)})^g}. \tag{4.3}$$

Nonlinear Least Square Fitting and Goodness of Fit

We use the *Nonlinear Least Squares* (NLS) (87) to estimate the parameters of the various models while fitting them with our data. Consider a set of n observations $(t_1, x_1), \ldots, (t_n, x_n)$ and a model function depending on m parameters $y = f(x, \boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$ and $n \geq m$. We want to find the vector $\boldsymbol{\beta}$ that minimizes the sum of squares:

$$S = \sum_{i=1}^{n} r_i^2,$$

where the residuals errors r_i are given by

$$r_i = y_i - f(x_i, \boldsymbol{\beta}),$$

for $i = 1, 2, \ldots, m$.

We tested the goodness of our fit by means of the *Kolmogorov-Smirnov Test* (KS) and *Maximum Likelihood Estimate* (MLE).

Advanced spectral analysis and trend extraction procedure

Singular-spectrum analysis (SSA) is a not-conventional spectral analysis method which provides insight into the unknown and/or partially known dynamics of a dynamical system (88; 89). More in detail, SSA aims at decomposing the signal as a linear combination of variability modes, which are data-adaptive functions of time. Thus, with respect to more traditional spectral approaches such as the classical Fourier decomposition, SSA doesn't ground on variability modes which have to be necessarily harmonic components. As a consequence, SSA provides a powerful de-noising filter, to identify the different components of the analyzed signal, such as trends, oscillatory patterns, harmonic and/or anharmonic oscillations, quasi-periodic phenomena, without making any assumption about the underlying generating of the observed signal (90). Moreover, SSA doesn't require the assumption of any particular stationarity or ergodicity conditions.

In order to distinguish between significant signal and random fluctuations (i.e. background noise), **Monte-Carlo (MCSSA)** is applied. MCSSA grounds on a particular Monte Carlo approach to the signal-to-noise separation issue, suited to overcome the limitations of classical signal extraction procedure, i.e. the identification of simply a gap in the eigenvalues spectrum (91). Recent fine-tunings of the method have been proposed to further improve results robustness and reliability in short time series (92). In the present work, MCSSA is applied to science and conspiracy community size time series, to establish whether our time series are linearly distinguishable from the linear stochastic processes, usually considered as noise. Both white and red noise null-hypotheses are taken

into account, since the choice of the most suitable kind of noise in social sciences, when dealing with advanced spectral methodologies, is still under debate.

Sentiment Classification

The sentiment classification is carried out as in (83) and refers to the same dataset. We consider three values for the sentiment of each comment: *negative* (-1), *neutral* (0), and *positive* (+1). We perform an automatic sentiment classification based on supervised machine learning that consists of the following four steps: (i) a sample of texts is manually annotated with sentiment (in our case 20K randomly selected comments are manually annotated by 22 native Italian speakers), (ii) the labeled set is used to train and tune a classifier, (iii) the classifier is evaluated on an independent test set or by cross-validation, and (iv) the classifier is applied to the whole set of texts. For more details on the classifier or on its performance refer to (83).

4.2 Community Evolution

Online social networks might elicit the aggregation of individuals in communities of interest. For the particular case of science and conspiracy users on the Italian Facebook, the emergence of two separate echo chambers has already been shown in a previous study (29). However little is known about the structural evolution of the two communities and the role of users' engagement in shaping them. To shade light on the determinants of group formation, as a first step, we analyze and compare the temporal evolution of science and conspiracy communities size by considering users commenting activity.

More in details, we divide users in three categories :

- U_1 the set of all active users – i.e. of all those users that commented at least once,
- U_2 the set of all users that commented at least twice, and

- U_5 the set of all users that commented at least five times.

For each set of users we look at the temporal evolution of the science and conspiracy communities, defined as:

$$S_i(t) = \left\{ u \in U_i : \frac{s_u}{n_u} \geq 0.95 \right\} \text{ and } C_i(t) = \left\{ u \in U_i : \frac{c_u}{n_u} \geq 0.95 \right\},$$

where $i \in \{1, 2, 5\}$, n_u is the total number of comments made by user u, s_u is the number of comments that user u made on science posts, c_u is the number of comments that user u made on conspiracy posts, and $t \in \{1, \ldots, T\}$[2]. We consider the threshold of 0.95 for the membership inside one community in accordance with previous studies (22; 65).

Figure 13 shows the temporal evolution of the size of the communities resulting from the previous classification. The dataset has been sampled by daily resolution, over the period January 2010–April 2012, for a total of 835-days observations. A similar global behavior emerges in all cases, and significant quantitative differences arise between C_1 (or C_2) and C_5, as well as between C_1 (or C_2) and S_i, $i \in \{1, 2, 5\}$. This phenomenon may be linked to the abundance of low-activity users inside the conspiracy communities, and for this reason in the next sections we will restrict our attention to the respective most active communities, S_5 and C_5. We also pairwise compared the six sample distributions by means of the Kolmogorov-Smirnov test (see Tab. 8 for the tests' results). For each users typology, we reject the null hypothesis of equivalence between science and conspiracy distributions, at the 99% confidence level.

In Fig. 14 we report the summarizing statistics for the users' mobility inside one particular community by box and whiskers plots (93) (or, simply boxplots). Black horizontal lines represent the median of the number of users entering or exiting the science and conspiracy communities, and the colored boxes represent the interquartile ranges (i.e., the 25th-75th percentile ranges) and they statistically measure the degree of dispersion and the skewness of each analyzed distribution: the users which

[2]The observation is carried out over the period January 2010–April 2012, by daily temporal resolution. T is the number of days of observation and it is equal to 835.

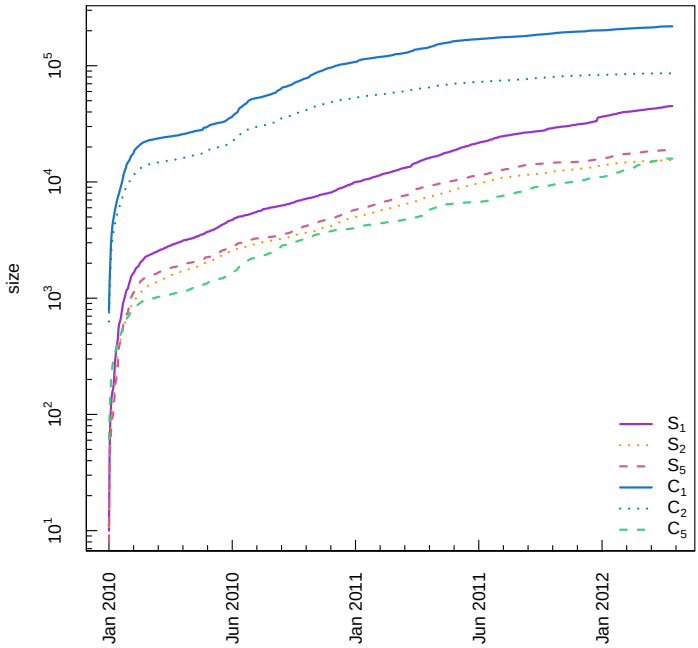


Figure 13: Temporal evolution of the size of the communities S_1 (solid violet), S_2 (dotted orange), S_5 (dashed pink), C_1 (solid blue), C_2 (dotted sea green), and C_5 (dashed green). The observation is carried out in the period January 2010–April 2012, with daily temporal resolution.

enter the science and conspiracy communities (violet and blue boxes, respectively), and the users which exit from each community (green and orange boxes for science and conspiracy, respectively). Vertical lines (i.e., the whiskers) are lower and upper bounded by the minimum and maximum values of the corresponding distribution, once both outliers and extreme values are removed from the data. Individual points represent the outliers of each analyzed distribution. From the left to the right, each set of boxplots corresponds to one user's category (i.e., U_1, U_2, and U_5).

In all cases we notice a significant difference between the users enter-

	D	C	p
$\mathbf{S_1/C_1}$	0.763	0.079	2.2×10^{-16}
$\mathbf{S_2/C_2}$	0.886	0.079	2.2×10^{-16}
$\mathbf{S_5/C_5}$	0.970	0.079	2.2×10^{-16}

Table 8: Results from Kolmogorov-Smirnov tests. D is the estimated maximum distance between the two distributions under analysis, C is the corresponding critical value, and p the resulting p-value. Considering a level of significance $\alpha = 0.01$, we reject the null hypothesis of equivalence of the two distributions in all the cases.

ing into and exiting from a community, favorable to the formers, indeed more than 99% of the users' flow is made up of those users entering one community.

These two results underline that the behavior of users is similar for both categories, irrespective of the contents. After an initial spike-like growth, the communities evolve at a nearly constant rate. Moreover, once a user enters one community the probability to get out of it is very small.

To better characterize the temporal evolution of both communities, we fit the Gompertz growth model (GM) in (4.1), the Logistic model (LM3, LM5) in (4.2), and the Log-logistic model (LLM) in (4.3) to our sample distributions S_5 and C_5, representing the temporal profile of quite active users, i.e. with at least 5 total comments, affiliated to science or conspiracy communities, respectively. The models are chosen on the basis of the observed evolution of the communities' size, that is characterized by a first phase of rapid growth, approximately exponential, followed by a more gradual one.

For each model we estimate its parameters through the *Nonlinear Least Squares NLS* (see Section *Methods* for more details about the fitting models). Fit's results are shown in Fig. 15 for both science (panel a) and conspiracy (panel b). Four fits are superposed to the original data (dashed green line): GM (solid orange line), LM3 (dotted violet line), LM5 (dot-dashed blue line), and LLM (dashed purple line).

As it can be deduced from Fig. 15, all models show a good approximation of the temporal evolution of science and conspiracy communi-

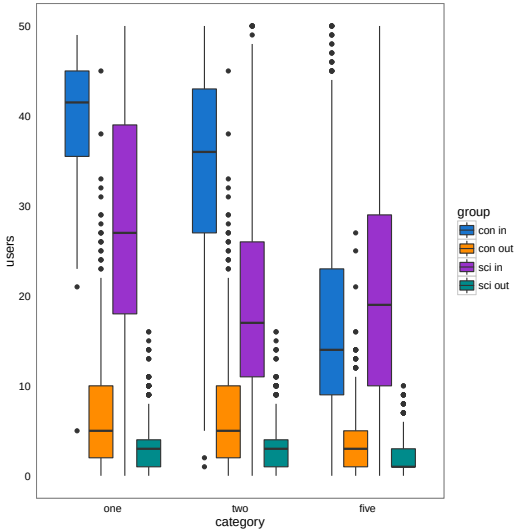


Figure 14: Boxplots of the mobility of users within each group. The black horizontal lines represent the median of the number of users entering or exiting the science and conspiracy communities for each temporal step. Individual points refer to the outliers of the distributions. Colored boxes represent the interquartile range (25th-75th percentile range), where blue stands for incoming users in conspiracy community, violet for incoming users in science, orange for exiting users from conspiracy, and green for exiting users from science. On the x-axis we have, from left to right, results for C_1, S_1, C_2, S_2, C_5, and S_5.

ties sizes. Anyway, in order to determine the quality of each fit and eventually identify the best one, we perform a series of Kolmogorov-Smirnov tests (KS) between the real data and each of the synthetic distributions and Maximum Likelihood Estimates (MLE) of each fit. Results of the KS tests are reported in Tab. 9. By considering a level of significance $\alpha = 0.01$, we fail to reject the null hypothesis of equivalence of the two distributions in all cases. The Logistic model maximizes the log-likelihood for both S_5 and C_5.

The particular *S*-shaped behavior observed on raw data, and then characterized by growth-model fits, reminds the one observed in the

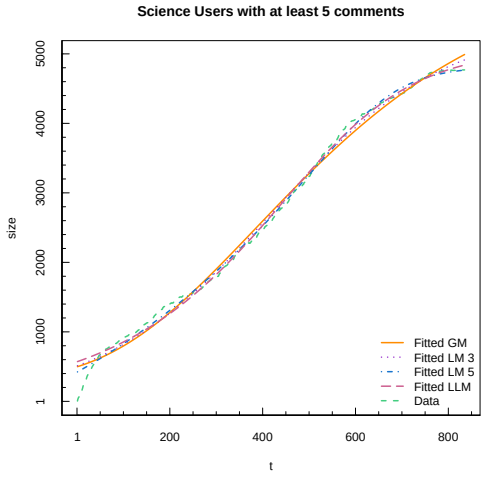
	D	C	p		D	C	p
$\mathbf{S_5/GM}$	0.071	0.079	0.031	$\mathbf{C_5/GM}$	0.060	0.079	0.100
$\mathbf{S_5/LM3}$	0.061	0.079	0.089	$\mathbf{C_5/LM3}$	0.077	0.079	0.015
$\mathbf{S_5/LM5}$	**0.035**	0.079	0.267	$\mathbf{C_5/LM5}$	**0.050**	0.079	0.241
$\mathbf{S_5/LLM}$	0.047	0.079	0.322	$\mathbf{C_5/LLM}$	0.053	0.079	0.197

Table 9: Results from Kolmogorov-Smirnov test. D is the estimated maximum distance between the two distributions under analysis, C is the corresponding critical value, and p the resulting p-value. Considering a level of significance $\alpha = 0.01$, we fail to reject the null hypothesis of equivalence of the two distributions in all cases.

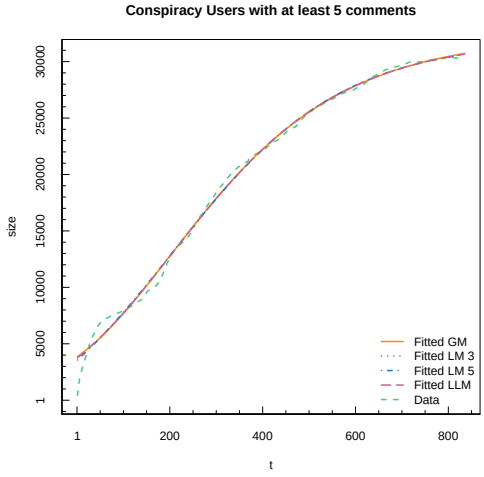
framework of population growth, where after a first stage of huge growth, a saturation level is reached, and population stabilizes. Logistic and Gompertz growth models found several fields of application, ranging from demography and sociology, to biology and ecology.

Science and conspiracy communities reach a thresholding value in their sizes growth, as fit results suggest. Those users which are deeply engaged in a community are more likely to become focused on a particular topic, and their increasing involvement into highly specified topics makes them "isolated" from the neighboring environment, which in this case is the whole world of knowledge. What is curious is that both conspiracy and science communities show the same size profiles.

To better assess the reliability of model fits results, we further inspect the time evolution of S_5 and C_5 communities sizes through advanced spectral methodologies extremely useful to uncover the presence of significant oscillatory movements, besides the huge growing trend dominating both communities growth. More precisely, we try to identify trends, oscillatory components (both periodic and not-periodic), and background noise in our series to finally reconstruct the embedded true signal, by summing up the contributions of all its significant components. We chose non-parametric methods, such as singular-spectrum analysis and similar methodologies (88; 89), in order to analyze our records time evolution by an alternative approach, which is not based on fitting an assumed model to the data, with the final goal in mind to further support model fits results by a completely different method.



(a)



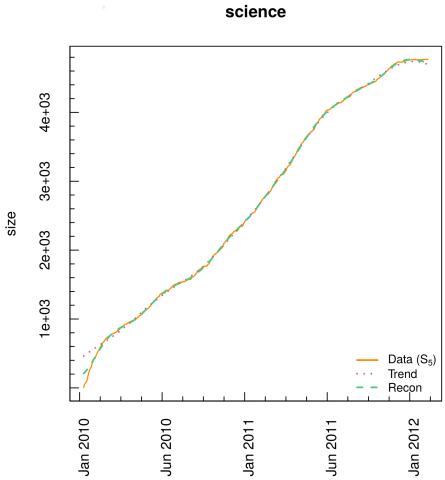
(b)

Figure 15: Fit of the temporal evolution of the size of Science (a) and Conspiracy (b) communities. We fitted the data (dashed green line) with four growth models: GM (solid orange), LM3 (dotted violet), LM5 (dot-dashed blue), and LLM (dashed purple). All model show a good fit for our data samples.

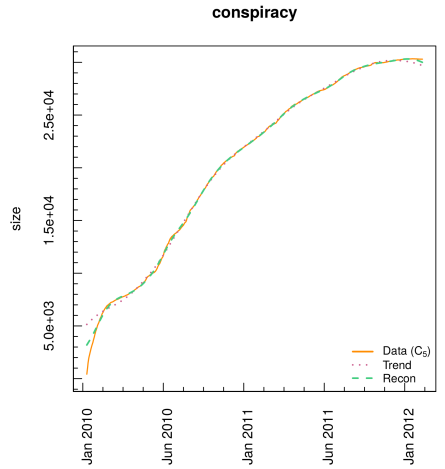
Indeed, the simultaneous and flexible application of more than one spectral tool can assure a quite reliable and robust analysis of temporal dynamics, especially when the signal-to-noise ratio is low, besides dealing with finite sample length. Moreover Monte-Carlo SSA (MC-SSA) (91; 92) test is applied to assess the significance of the revealed oscillatory modes with respect to both white and red noise background noise null-hypotheses. The reader is referred to Section *Methods* for deeper details about the applied methodology.

Both conspiracy and science time series behavior turn out to be described by the first two T-PCs (temporal principal components), which in that case correspond to the trend. More in details, the trends capture the 96.16% and the 95.44% of S_5 and C_5 series total variance, respectively. Besides, we extracted the pure significant reconstructed signals from our series, and we observed that they turned out to be quite similar to trends (exception made for some boundary effects due to the finite-sample length). Figure 16 shows the trends (dotted violet line) and reconstructed signals (dashed green line) superposed to S_5 (top) and C_5 (bottom) communities size evolution in time (orange lines). Boundary effects are visible, especially at the beginning of the series, but quite negligible. Trends are able to catch both S_5 and C_5 temporal profile, and they mainly coincide with the reconstructed significant signals, in both cases.

As a further check, we pre-process data, first by removing the trend, second by standardizing-by-trends the so obtained residual time series. Pre-processing is required since the presence of such a pervasive trend reflects in a high peak at zero frequency dominating the shape of power spectrum estimate, and sometimes hiding eventual higher-frequency cycles. No significant cycle is detected in S_5 and C_5 series after trend removal. Figure 17 shows S_5 and C_5 detrended time series (panel a) and S_5 and C_5 residual time series standardized by their trends (panel b). The apparent oscillating behavior visible in raw data and in the detrended time series (especially in C_5, Fig. 17(a)) is not connected to significant oscillatory modes, according to Monte-Carlo SSA test. Besides, both communities show a smoother profile after Jan 2011 (Fig. 17(b)), in the range $t > 600$ in Fig. 15. At that time, both S_5 and C_5 growth starts to decrease.

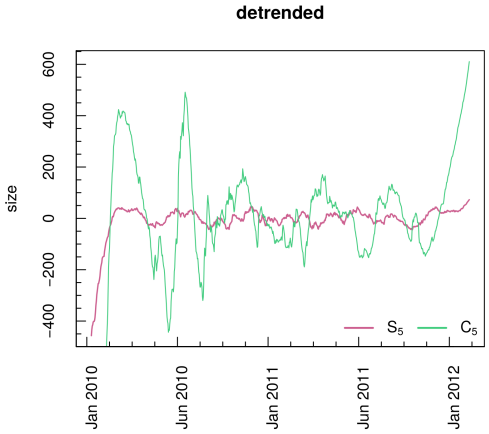


(a)

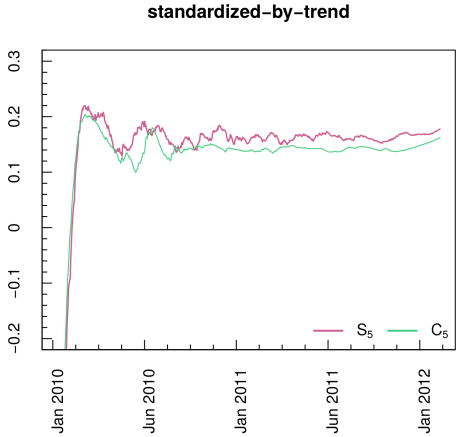


(b)

Figure 16: S_5 (a) and C_5 (b) dominant spectral components. Original series are shown in solid orange lines, trends in dotted violet lines, and significant signal reconstructions in dashed green lines. A pervasive trend dominates both Science and Conspiracy communities sizes temporal evolution.



(a)



(b)

Figure 17: Pre-processing procedure. (a) Detrended S_5 (solid pink) and C_5 (solid green). (b) Standardized-by-trend S_5 (solid pink) and C_5 (solid green) residual time series. In panel b, the pre-processed series are standardized to zero mean and unit variance.

Notice that both the pre-processed time series shown in Fig. 17(b) are standardized to zero mean and unit variance, in order to help the visual comparison between S_5 and C_5. We can finally infer that the trends determine the time evolution of our records, only. Thus, we compare the previous described model fits (GM, LM3, LM5, LLM) to the S_5 and C_5 trends, only. No particular difference emerges between science and conspiracy communities in terms of their growth, and the linear correlation between both communities trends and each fitted model turns out to be very high for all the cases, preventing us to identify a significantly favorite fit, in agreement with the results previously reported in Tab. 9. *Pearson* correlation coefficient is computed, since no particular significant cycle emerges from S_5 and C_5 sizes records spectral analysis, thus reducing the risk of underestimating the presence of an eventual correlation at time-shifted version of the original series.

Our analysis thus suggests that communities present strong similarities, and that the behavior of users inside each of them is similar. Once they have selected their preferred group, users seem to undergo community dynamics, that are similar in both science and conspiracy case, irrespectively of the content.

4.3 Users' Sentiment Analysis

Now we zoom in at the level of the emotional dynamics of the polarized groups. We approximate the emotional attitude of users towards one piece of information that they commented by considering the sentiment of the text. We label the sentiment of each comment as: *negative* (-1), *neutral* (0), or *positive* (+1). We perform an automatic sentiment classification based on supervised machine learning, refer to Section *Materials and Methods* or to (83) for more details.

Our aim is to characterize the emotional behavior of the users as a function of their involvement inside the community. To do this we define three new measures, the *mean user sentiment* (σ_i), the *mean negative/positive difference of comments* ($\delta_{NP}(i)$), and the *user sentiment polariza-*

tion ($\varrho_\sigma(i)$) as it follows:

$$\delta_{NP}(i) = \frac{1}{T_i}\sum_{j=1}^{T_i}(Neg_j(i) - Pos_j(i)), \tag{4.4}$$

where T_i is the number of days in which user i was active, $Neg_j(i)$ the number of i's negative comments in day j, $Pos_j(i)$ the number of i's positive comments in day j;

$$\varrho_\sigma(i) = \frac{(N_i - 2k_i - h_i)(N_i - h_i)}{N_i^2}, \tag{4.5}$$

where N_i, k_i, h_i are respectively the number of all, negative, and neutral comments left by user i, while $l_i = N_i - k_i - h_i$ is the number of the positive ones. Note that $\varrho_\sigma(i) \in [-1, 1]$ and that it is equal to 0 if and only if $l_i = k_i$ or $h_i = N_i$, it is equal to 1 if and only if $k_i = N_i$, and it is equal to -1 if and only if $l_i = N_i$. Finally, σ_i is simply defined as the mean of the sentiment of all comments left by user i.

Figure 18 shows the average sentiment σ_i for science users (panel a), conspiracy users (panel b), and all users (panel c), as a function of the user engagement – i.e., the total number of comments left by each user. In the insets we report, for each of the three categories, the value of σ_i as a function of the number of comments for the most active users, i.e. those users with at least 100 comments. We then regress the mean user sentiment σ_i w.r.t. the logarithm of the number of comments. We notice that σ_i becomes more negative as the number of comments increases, in all cases, when considering all users. However, when we restrict our attention to the most active users, we notice that σ_i becomes more negative as the number of comments increases only in science case, while the opposite is true for the other cases.

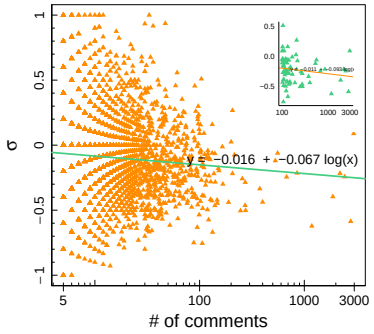
Figure 19 shows the mean negative/positive difference of comments $\delta_{NP}(i)$ of science users (panel a), conspiracy users (panel b), and all users (panel a), as a function of the user engagement. In the insets we report, for each of the three categories, the value of $\delta_{NP}(i)$ as a function of the number of comments for those users with at least 100 comments. We regress the mean negative/positive difference $\delta_{NP}(i)$ w.r.t. the logarithm

of the number of comments. $\delta_{NP}(i)$ is a measure of the mean negative shift from a situation of neutral equilibrium for which either the user has only neutral comments or she has the same number of positive and negative comments. A positive value of $\delta_{NP}(i)$ indicates that the user tends to have, on average, more negative than positive comments. From Fig. 19 we notice that $\delta_{NP}(i)$ tends to increase when the number of comments increases in all cases, underlining the fact that, on average, more active users tend to show a faster shift towards the negativity than less active ones. The rate of this increment in the negativity is higher for users with more than 100 comments and it is also higher for science users w.r.t conspiracy ones.

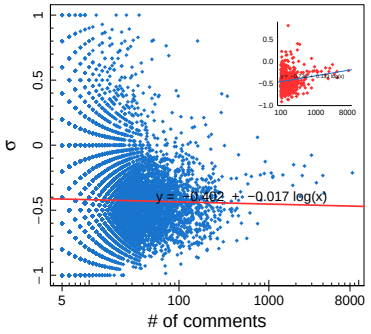
While Figure 20 reports the user sentiment polarization $\varrho_\sigma(i)$ of science users (panel a), conspiracy users (panel b), and all users (panel c), as a function of the user engagement. In the insets we show, for each of the three categories, the value of $\varrho_\sigma(i)$ as a function of the number of comments for those users with at least 100 comments. We regress the user sentiment polarization $\varrho_\sigma(i)$ w.r.t. the logarithm of the number of comments. The user sentiment polarization $\varrho_\sigma(i)$ ranges in $[-1, 1]$, and it is equal to 0 either if all comments are neutral or if there is the same number of negative and positive comments, while it tends to 1 (resp. -1) when $l_i \gg k_i$ and h_i is small enough, i.e., when the number of positive comments is much bigger than the number of negative ones, (resp. $k_i \gg l_i$ and h_i is small enough, i.e., when the number of negative comments is much bigger than the number of positive ones). Science users show an higher value of $\varrho_\sigma(i)$, however conspiracy users with at least 100 total comments tend to increase it w.r.t. science ones.

The engagement within the echo chamber affects users emotional dynamics. The more a user is active, the higher the tendency to express negative emotion when commenting. This feature holds for both users categories. Moreover, for both categories we observe that, on average, more active users show a faster shift towards the negativity than less active ones. The rate of this increment in the negativity is higher for users with more than 100 comments and it is also higher for science users w.r.t conspiracy ones. In terms of the users' sentiment polarization we ob-

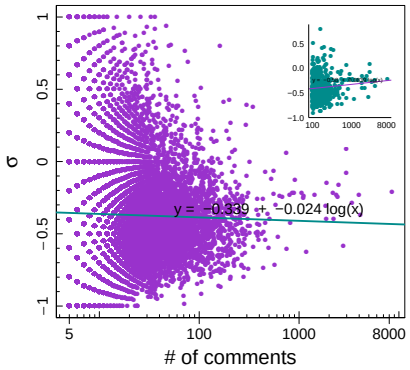
serve some differences between the two categories: its value is generally higher for science users, however very active science users tend to decrease their sentiment polarization with the increasing of the activity, while on the contrary conspiracy ones tend to increase it.



(a)

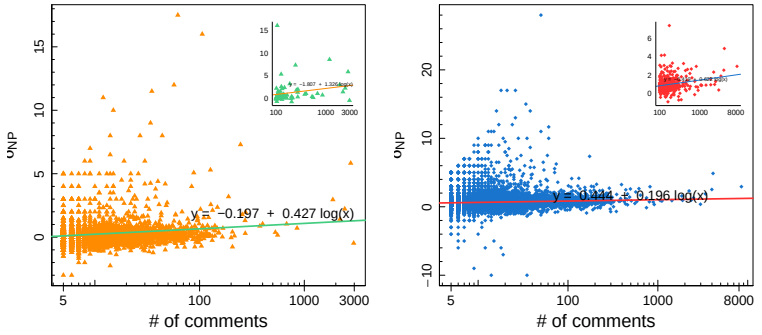


(b)

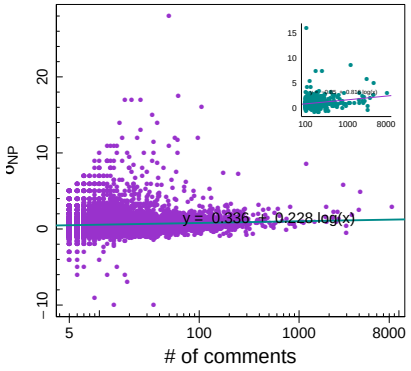


(c)

Figure 18: Mean final sentiment σ_i of science users (a), conspiracy users (b), and all users (c), as a function of the user engagement i.e., the total number of comments left by each user. In the insets we report, for each of the three categories, the value of σ_i as a function of the number of comments for those users with at least 100 comments. We regress the mean user sentiment σ_i w.r.t. the logarithm of the number of comments.

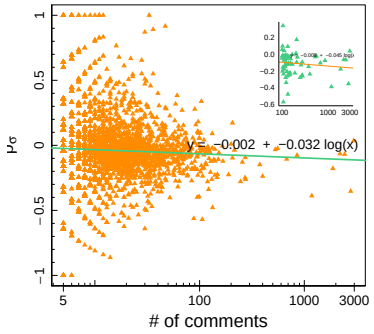


(a) (b)

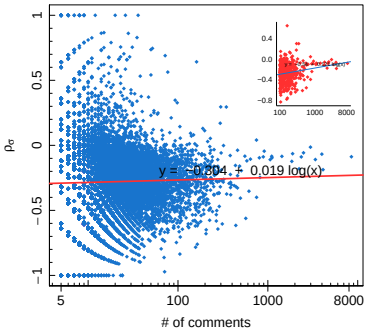


(c)

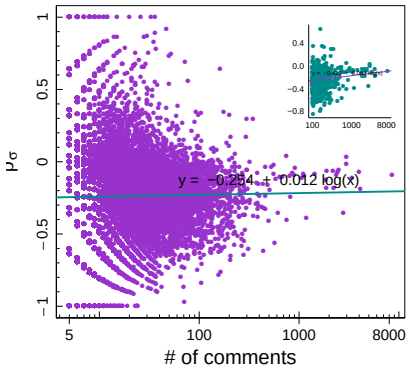
Figure 19: Mean negative/positive difference $\delta_{NP}(i)$ of science users (a), conspiracy users (b), and all users (c), as a function of the user engagement. In the insets we report, for each of the three categories, the value of $\delta_{NP}(i)$ as a function of the number of comments for those users with at least 100 comments. We regress the mean negative/positive difference $\delta_{NP}(i)$ w.r.t. the logarithm of the number of comments.



(a)



(b)



(c)

Figure 20: User's sentiment polarization $\varrho_\sigma(i)$ of science users (a), conspiracy users (b), and all users (c), as a function of the user engagement. In the insets we report, for each of the three categories, the value of $\varrho_\sigma(i)$ as a function of the number of comments for those users with at least 100 comments. We regress the user sentiment polarization $\varrho_\sigma(i)$ w.r.t. the logarithm of the number of comments.

4.4 Evolution of the Sentiment inside the Communities

We now focus on the collective sentiment of the two communities, rather than the single user's one. Similarly to the single user case we define the *community negative/positive difference* of comments (δ_{NP}^{C}) and the *mean community sentiment polarization* (ϱ_{σ}^{C}) as follows:

$$\delta_{NP}^{C} = \frac{1}{M_C}\left(\frac{1}{T}\sum_{j=1}^{T}(Neg_j^C - Pos_j^C)\right), \tag{4.6}$$

where T is the number of days of observation, Neg_j^C the number of negative comments from users belonging to community C during day j, Pos_j^C the number of positive comments from users belonging to community C during day j, M_C is the maximum daily activity of community C, and $C \in \{Science, Conspiracy\}$; While

$$\varrho_{\sigma}^{C} = \frac{(N_C - 2k_C - h_C)(N_C - h_C)}{N_C^2}, \tag{4.7}$$

where N_C, k_C, h_C are respectively the number of all, negative, and neutral comments left by users of community C, while $l_C = N_C - k_C - h_C$ is the number of positive ones. Note that $\varrho_{\sigma}^{C} \in [-1, 1]$.

Figure 21 displays the community negative/positive difference of comments δ_{NP}^{C} as a function of the daily community activity for science users (left) and conspiracy users (right). The top figures show the values considering all users in the communities, while the bottom ones only consider those users with at least 100 comments. We regress the community negative/positive difference of comments δ_{NP}^{C} w.r.t. the logarithm of the number of comments inside the community at a given time. For both communities δ_{NP}^{C} tend to increase; while science one shows an higher increasing rate for the most active case, conspiracy one shows an higher increasing rate for the general case.

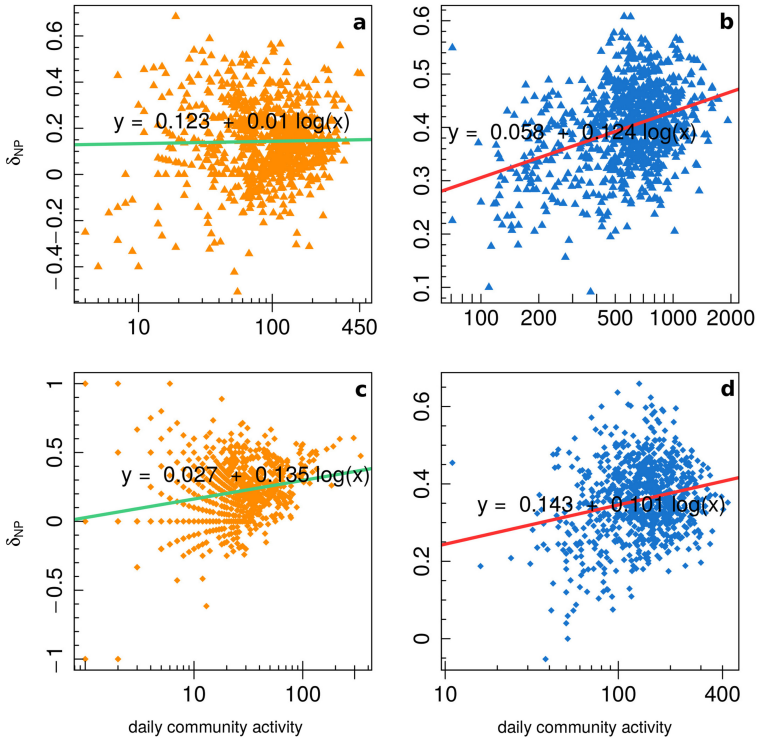


Figure 21: Community negative/positive difference of comments δ^C_{NP} as a function of the daily community activity for science users (left) and conspiracy users (right). The top figures show the values of δ^C_{NP} considering all users in the communities, while the bottom ones only consider those users with at least 100 comments. We regress the community negative/positive difference of comments δ^C_{NP} w.r.t. the logarithm of the number of comments inside the community at a given time.

Figure 22 shows the mean community sentiment polarization ϱ_σ^C as a function of the daily community activity for science users (left) and conspiracy users (right). The top figures display the values considering all users in the communities, while the bottom ones only consider those users with at least 100 comments. We regress the mean community sentiment polarization ϱ_σ^C w.r.t. the logarithm of the number of comments inside the community at a given time. For the conspiracy community we notice a decrement in the value of ϱ_σ^C as the number of comments increases, moreover this decrement is higher for most active users. Science community instead shows a decrement in the value of ϱ_σ^C for the case of most active users and a slight increment for the general case.

Also the community sentiment behavior is affected by the cumulative users' activity (in terms of comments). When either community is more active, the shift towards negative comments is larger. A difference between the two echo chambers comes upon if we restrict our attention only to the most active users, i.e. those with at least 100 comments. In this last case, science users show a higher rate of increment than conspiracy ones, contrary to the general case. Differently from the single user case, the community sentiment polarization shows a deep decrement with higher activity in the conspiracy community, the process is slower for science community, when we consider only most active users, and even reversed in the general case.

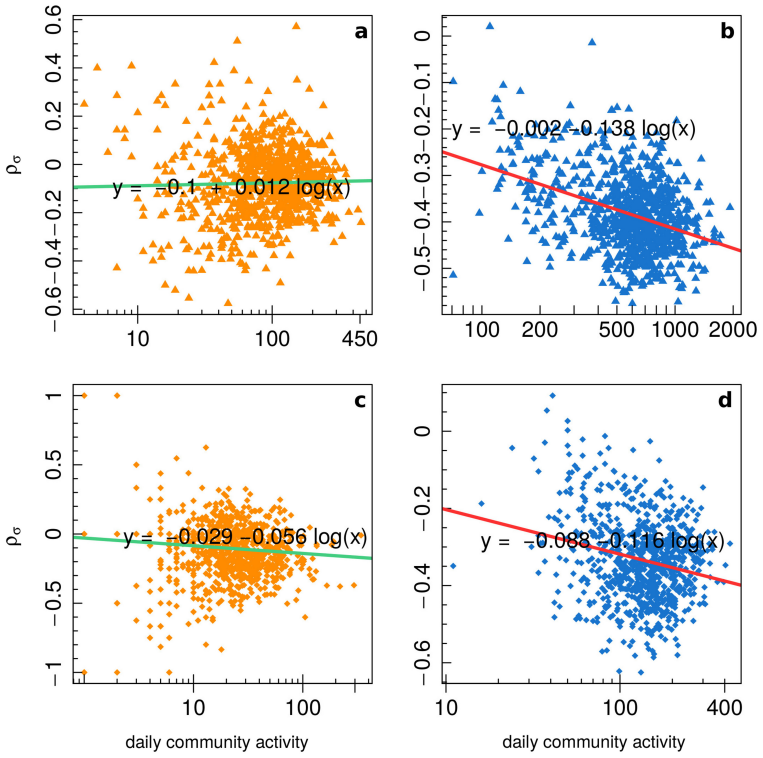


Figure 22: Mean community sentiment polarization ϱ_σ^C as a function of the daily community activity for science users (left) and conspiracy users (right). The top figures show the values of ϱ_σ^C considering all users in the communities, while the bottom ones only consider those users with at least 100 comments. We regress the community negative/positive difference of comments ϱ_σ^C w.r.t. the logarithm of the number of comments inside the community at a given time.

4.5 Concluding Remarks

The Facebook environment is particularly suited for the emergence of polarized communities, or echo chambers. The activity inside such echo chambers is limited to only one type of content. In this chapter, we report results on the behavior of users inside the echo chamber and on the structural evolution of the community accounting for both users activity and the sentiment they express (30).

We first study the evolution of the size of the two communities by fitting daily resolution data with three growth models, i.e. the Gompertz model, the Logistic model, and the Log-logistic model, and we observe that both communities evolve in a similar way and the behavior of users is similar irrespectively of the difference in contents: after a first phase of rapid growth, approximately exponential, both the communities sizes undergo a more gradual growth, until a thresholding value is reached. This behavior reminds malignant cancer evolution dynamics, where after a huge proliferation of anomalous cells, the lack of a sufficient vascularization inside central malignant core limits the very own growth. In our case, the lack of communication with the environment can be supposed to associate with the users extreme focusing on a precise topic, which alienates them from the real knowledge world.

Then we notice that both the users' and the communities' emotional behavior is affected by the users' involvement inside the echo chamber. To an higher involvement corresponds a more negative approach. Moreover, for both categories we observe that, on average, more active users show a faster shift towards the negativity than less active ones. The rate of this increment in the negativity is higher for users with more than 100 comments and it is also higher for science users w.r.t conspiracy ones. The community sentiment polarization shows a deeper decrement with higher activity in the conspiracy community, while the process is slower for science community, when we consider only most active users, and even reversed in the general case.

Chapter 5

Modeling Opinion Dynamics on Networks

The role of Confirmation Bias and Polarization in Opinions Formation

All the results shown in this chapter refer to the submitted paper (31)[1]. To better understand the role of confirmation bias and social influence in selecting the information and fostering the aggregation of online users in polarized groups, we provide a mathematical model mimicking polarization in online social dynamics.

Online users tend to select claims that adhere to their system of beliefs and to ignore dissenting information (64; 65; 84; 94; 95). The wide availability of content on the web fosters the aggregation of likeminded people where the discussion tends to enforce group polarization (82; 83). Confirmation bias, indeed, plays a pivotal role in viral phenomena (29). Under such conditions public debates, in particular on social relevant

[1]The results shown in this Chapter are all part of the paper (31), available as a pre-print at arXiv:1509.00189. It is a joint work with Dr. Antonio Scala, Prof. Guido Caldarelli, Prof. H. Eugene Stanley, and Dr. Walter Quattrociocchi. MDV, AS, and WQ outlined the research question. MDV performed the simulations. MDV and AS interpreted the results. MDV, AS, GC, HES, and WQ contributed equally to the writing and reviewing of the manuscript.

issues, tend to further fragment and polarize the public opinion (96; 97).

Opinion dynamics, have been widely investigated in recent years, using different approaches from statistical physics and network science (3). Classical examples of opinion dynamics models include the Sznajd model (39), the voter model (34; 35; 98), the majority rule model (38; 99), and the bounded confidence model (BCM) (5; 40; 100). Besides the different assumptions and dynamical rules, for all the cited models the consensus state, in which all agents share the same opinion, is reached under the right conditions.

However, consensus in far from common in real world and Internet based opinion exchanges. A recent study showed the emergence of *polarized communities*, i.e., *echo chambers*, in online social networks (29). Inside these communities, homogeneity appears to be the primary driver for the diffusion of contents. Both polarization and homogeneity might be the result of the conjugate effect of *confirmation bias* and *social influence*. Confirmation bias is the tendency to acquire or process new information in a way that confirms one's preconceptions and avoids contradiction with prior belief (23). Social influence is the process under which one's emotions, opinions, or behaviors are affected by others. In particular, *informational influence* occurs when individuals accept information from others as evidence about reality (11; 17).

Previous studies (8; 44) proposed a non consensus opinion model (NCO) that allowed for the stable coexistence of two opinions by also considering the opinion of the user herself when applying the majority rule update (8), while in (44) the competition between two groups is investigated by the introduction of a set of contrarians in one of the two. The survival of a two-opinions state is studied in (14) from a different point of view, considering the emergence of spontaneous recovery of failed nodes and the majority rule update. Both these models assume only two opinion states (± 1) and a majority rule update, with the novelty of accounting for the individual opinion (8; 44) and for an external source of influence (14).

Authors in (101) investigate the emergence of extreme opinion trends in society by employing statistical physics modeling and analysis on

polls. By developing an activation model of opinion dynamics with interaction rules based on the existence of individual stubbornness, they discover a sharp statistical predictor of the rise of extreme opinion trends in society in terms of a nonlinear behavior of the number of individuals holding a certain extreme view and the number of individuals with a moderate opinion and extreme opinion. A model grounded on the BCM and accounting for the interconnection and complexity of the online environment as well as the competition among sources of information is presented in (80). In a recent study(102), authors analyze the effects of the interplay between homophily, social influence, and confirmation bias in the emergence of segregation and echo chambers.

People shape their opinions on the basis of both confirmation bias and social influence, a combination of these two forces generates the observed polarization of communities and homogeneous links (29). Accounting for this phenomenon, we build a model of opinion dynamics and network's evolution that considers both mechanisms and expands itself from the classical *Bounded Confidence Model* (BCM) (5). We consider two variations of the model: the *Rewire with Bounded Confidence Model* (RBCM), in which discordant links are broken until convergence is reached; and the *Unbounded Confidence Model*, under which interaction among discordant pairs of users is allowed and a negative updating rule is introduced, either with the rewiring step (RUCM) or without it (UCM). As for the BCM, our models assume a continuous interval of opinions.

The chapter is structured as it follows. In the first section we provide references to the methods employed and give a brief overview of the BCM and its convergence results. In the *Results and Discussion* section, we first present the new models and give an account of the simulation results, then we present a mean field approximation of the newly introduced models.

Materials and Methods

Periodic Boundary Conditions

We consider N agents and a set of initial opinions x_i, $i \in \{1, \ldots, N\}$, uniformly distributed in $[0, 1]$. If we compare two agents' opinions by the absolute value distance $|x_i - x_j|$, those agents with near boundary opinions will have less concordant peers by definitions. We can overcome this problem by using the *Periodic Boundary Conditions* (PBC) and the alternative opinions' distance $|.|_\tau : [0, 1] \times [0, 1] \to [0, 0.5]$ defined as:

$$|x_i - x_j|_\tau = |x_i - x_j - \rho(x_i - x_j)|,$$

for $i, j \in \{1 \ldots, N\}$. The $\rho(.) : [-1, 1] \to \{-1, 0, 1\}$ adjustment ensures PBC and it is defined as:

$$\rho(x) = \begin{cases} -1, & \text{if } x \in [-1, -0.5) \\ 0, & \text{if } x \in [-0.5, 0.5] \\ 1, & \text{if } x \in (0.5, 1] \end{cases} . \tag{5.1}$$

5.1 The Bounded Confidence Model (BCM)

The *Bounded Confidence Model* (BCM) (5; 100) considers a set of N agents arranged on a complex network G.[2] Each agent holds an opinion x_i, $i \in \{1, \ldots, N\}$, uniformly distributed in $[0, 1]$. Two agents interact if and only if they are connected in G and their present opinions are close enough, i.e. iff $j \in N_G(i)$ and $|x_i - x_j| < \varepsilon$, for $\varepsilon \in [0, 1]$.[3] If these conditions hold, the two agents change their opinions according to Eq. (5.2), otherwise

[2]We consider different types of complex networks in the simulations: The Erdös-Rényi random network (ER) (103) characterized by a Poisson degree distribution with average degree $\langle 2 \rangle$, the scale-free network (SF) (104) characterized by a power-law degree distribution $P(k) \sim k^{-\gamma}$ with $2 \leq \gamma \leq 3$, and the small-world network (SW) (69) with rewiring probability equal to 0.2 and neighborhood dimension equal to 2. We notice that the network structure does not influence the simulation results, for this reason and considering that the SF network is the one that better reproduce the structure of online social media, we restrict our attention to SF networks.

[3]We apply periodic boundary conditions in the simulations and hence two users will interact if: $|x_i - x_j|_\tau < \varepsilon$, for $\varepsilon \in [0, 0.5]$.

they do not interact at all:

$$\begin{cases} x_i = x_i + \mu(x_j - x_i) \\ x_j = x_j + \mu(x_i - x_j) \end{cases} , \tag{5.2}$$

where the convergence parameter μ is taken in the interval $(0, 0.5)$.

It is known from previous studies (41; 42) that for ε big enough consensus is reached. The time rate change of $\mathbb{P}(x,t)dx$, the fraction of agents whose opinion at time t lies in the interval $[x, x+dx]$, is given by:

$$\begin{aligned} \frac{\partial \mathbb{P}(x,t)}{\partial t} = & - \mathbb{P}(x,t) \int_{-\varepsilon}^{\varepsilon} \mathbb{P}(x+y,t)dy \\ & + \frac{1}{(1-\mu)} \int_{-\varepsilon-2x}^{\varepsilon-2x} \mathbb{P}(x+y,t)\mathbb{P}(x - \frac{\mu}{1-\mu}y, t)dy. \end{aligned}$$

The first two moments are given by $M_0 = \int \mathbb{P}(x,t)dx = 1$ and $M_1 = \int x\mathbb{P}(x,t)dx = 0$, i.e. the total mass and the mean opinion, and they are conserved (41). Let $\mathbb{P}(x,0) = 1$ be a flat initial condition, with $x \in [0,1]$. We are interested in the final state of the system $\mathbb{P}(x,\infty)$.

When all agents interact, i.e., when $\varepsilon \geq 1$ the rate equation is integrable.[4] The second moment obeys $\dot{M}_2 + M_0 M_2/2 = M_1^2$, and using $M_1 = 0$ and $M_0 = 1$ we find that $M_2(t) = M_2(0)e^{-t/2}$, hence the second moment vanishes exponentially in time, all agents approach the center opinion and the system eventually reaches consensus (41):

$$\mathbb{P}_\infty(x) = M_0 \delta(x).$$

When $\varepsilon \geq 1$ the final state is a single peak located in the middle and, as long as $\varepsilon \geq 1/2$, this situation persists.[5] For smaller values of the threshold ε, it has been shown, by numerical simulations, that consensus is not reached and the opinion evolves into clusters that are separated by a distance larger than ε. Once each cluster is isolated it evolves into a Dirac delta function as in the case $\varepsilon \geq 1$. The final distribution consists

[4] As we use PBC, $\varepsilon \geq 1/2$ in the simulations.

[5] Again, thanks to the PBC we get $\varepsilon \geq 1/4$ in the simulations.

of a series of non interacting clusters at locations x_i with masses m_i:

$$\mathbb{P}_\infty(x) = \sum_{i=1}^{r} m_i \delta(x - x_i),$$

where r is the number of evolving opinion clusters (41). All clusters must fulfill the conservation laws $\sum m_i = M_0 = 1$, and $\sum x_i m_i = M_1 = 0$ is equal to the conserved mean opinion. All different clusters $i \neq j$ must also fulfill $|x_i - x_j| > \varepsilon$.

5.2 Results and Discussion

5.2.1 Models

The paper is a model study derived from the paper (29) on which we provide evidence of the polarizing effect of different narratives and the echo chamber structure of cascades. Hence, here we exploit the bounded confidence proviso (i.e., interacting with an information/opinion iff this is close enough to the agent state) that well mimics the confirmation bias (i.e., acquiring information that adhere to a specific system of beliefs) process.

Starting from the BCM we introduce three new models of opinion dynamics and network evolution. The first model we consider is the **Rewire with Bounded Confidence Model** (RBCM) that considers the same framework as in BCM and involves two phases. In phase one we run the *rewiring steps* in which each agent i interacts with a randomly chosen neighbor j and, if the distance between the two opinions is above the tolerance ε, i.e. if $|x_i - x_j|_\tau \geq \varepsilon$, for $\varepsilon \in [0, 0.5]$,[6] then their link is broken and i is rewired to a randomly chosen agent $k \in \{1, \ldots, N\}/(N_G(i) \cup \{i\})$. Phase one ends when all links have an opinion distance below the tolerance ε. In phase two we run the BCM on the rewired network. As all the couples in the new network are agreeing, all the randomly chosen pairs will interact and readjust their opinion according to Eq. (5.2).

[6] We restrict our attention to $\varepsilon \in [0, 0.5]$ after noticing that $\forall x, y \in \{1, \ldots, N\}$ we get $|x - y|_\tau \in [0, 0.5]$. We will assume $\varepsilon \in [0, 0.5]$ throughout the paper.

The **Unbounded Confidence Model** (UCM) is the second of the models that we introduce and its novelty is to allow the interaction for every randomly chosen pair of neighbors (i, j). To be specific, if two agents agree, i.e. if $|x_i - x_j|_\tau < \varepsilon$, as for the previous model, we adjust x_i and x_j by rule (5.2). However, if the distance between their opinions is above the tolerance, i.e. if $|x_i - x_j|_\tau \geq \varepsilon$, we use a new updating rule in Eq. (5.3) that enables us to replicate the empirically observed repulsion of disagreeing opinions:

$$\begin{cases} x_i = x_i - \mu[x_j - x_i - \rho(x_j - x_i)] \\ x_j = x_j - \mu[x_i - x_j - \rho(x_i - x_j)] \end{cases}, \tag{5.3}$$

where the convergence parameter μ is taken in the interval $(0, 0.5)$ and $\rho(.)$ is defined in Eq. (5.1). The adjustment $\rho(.)$ ensures the PBC by maintaining the opinions inside the interval $[0, 1]$. The last model that we introduce is the **Rewire with Unbounded Confidence Model** (RUCM) that again allows the interaction for every randomly chosen pair of users (i, j) but at the same time allows for the random rewiring of disagreeing pairs. Specifically, if $|x_i - x_j|_\tau < \varepsilon$, then we adjust x_i and x_j by rule (5.2). If $|x_i - x_j|_\tau \geq \varepsilon$, then we adjust x_i and x_j by rule (5.3), the link between nodes i and j is broken, and a new link between i and a randomly chosen user $k \in \{1, \ldots, N\} \backslash (N_G(i) \cup \{i, j\})$ is created.

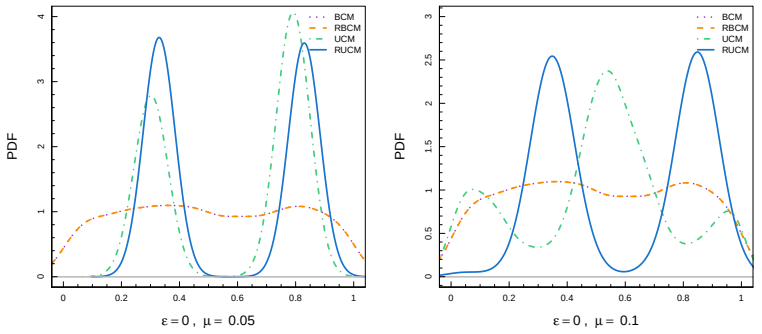
5.2.2 Simulation Results

We consider different types of complex networks in the simulations: The Erdös-Rényi random network (ER) (103) characterized by a Poisson degree distribution with average degree $\langle 2 \rangle$, the scale-free network (SF) (104) characterized by a power-law degree distribution $P(k) \sim k^{-\gamma}$ with $2 \leq \gamma \leq 3$[7], and the small-world network (SW) (69) with rewiring probability equal to 0.2 and neighborhood dimension equal to 2. Here we report results for the SF networks and while for those of the Erdös-Rényi random network and the small-word network refer to Fig. 30, 31.

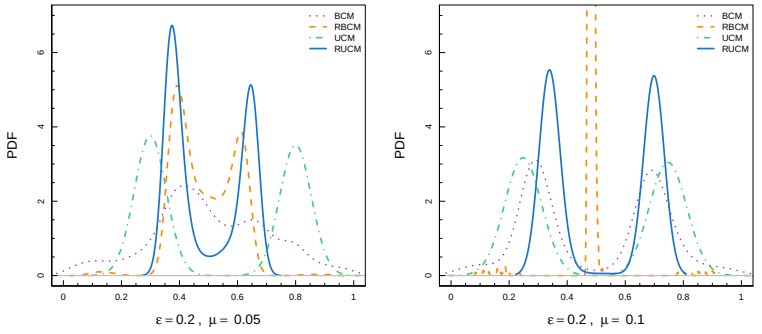
[7] The scale-free networks are created by using the classic implementation of the Barabási and Albert model.

Hence, we show the results of Monte Carlo simulations of the BCM and the three new models on a SF network of 2000 nodes with the parameters (ε, μ) varying in the parameter space $[0, 0.5] \times [0, 0.5]$, for $T = 10^5$ time steps and we averaged our results over 5 repetitions. Note that the final state, under the different parameters combinations, is always reached before $T = 10^5$. Refer to Fig .27 for further details. Figure 23 shows the probability density functions (PDFs) of final opinion, after a maximum of 10^5 time steps, for four different combinations of the pair of parameters (ε, μ): $(\varepsilon, \mu) \in \{(0, 0.05), (0, 0.1), (0.2, 0.05), (0.2, 0.1)\}$. The blue solid and the green dot-dashed curves refer to the newly introduced RUCM and UCM respectively, while the violet dotted curve is for BCM and the orange dashed for RBCM. For all the parameter choices we observe a bimodal opinion distribution in the cases of RUCM and UCM (note that we assume periodic boundary conditions). It is interesting to note that for UCM and RUCM there are two polarized opinions also for $\varepsilon = 0$, while in that case BCM and RBCM show no changes with respect to the initial uniform distribution.

Figure 24 reports a collection of summary statistics (mean, standard deviation, 1st quantile, and 3rd quantile) of the final opinion distributions for varying ε and three different values of μ (violet is for $\mu = 0.05$, blue for $\mu = 0.25$, and orange for $\mu = 0.5$). The left column is for BCM, the central one for UCM, and the right one for RUCM. We omit the results for RBCM as we observe from the simulations that, after the rewiring steps, the dynamics are similar to the BCM case but with a faster convergence, refer to the *Supplementary Information* online for an in depth analysis of the RBCM model. We observe different mechanisms for the two newly introduced models, such as a faster convergence to the consensus state for RUCM. However, we need to study the final number of peaks to better characterize the differences between UCM and RUCM, and to relate them with the results for the classical BCM.



(a) (b)



(c) (d)

Figure 23: Probability density functions (PDFs) of final opinion, after a maximum of 10^5 time steps or until convergence is reached, for four different combinations of the parameters (ε, μ). In panel (a) we have $(\varepsilon, \mu) = (0, 0.05)$, in panel (b) $(\varepsilon, \mu) = (0, 0.1)$, in panel (c) $(\varepsilon, \mu) = (0.2, 0.05)$, and in panel (d) $(\varepsilon, \mu) = (0.2, 0.1)$. In all panels the blue solid curve is for RUCM, the green dot-dashed one for UCM, the violet dotted one for BCM, and the pale orange dashed one for RBCM. We observe a bimodal distribution for RUCM and UCM, representing the coexistence of two polarized stable opinions.

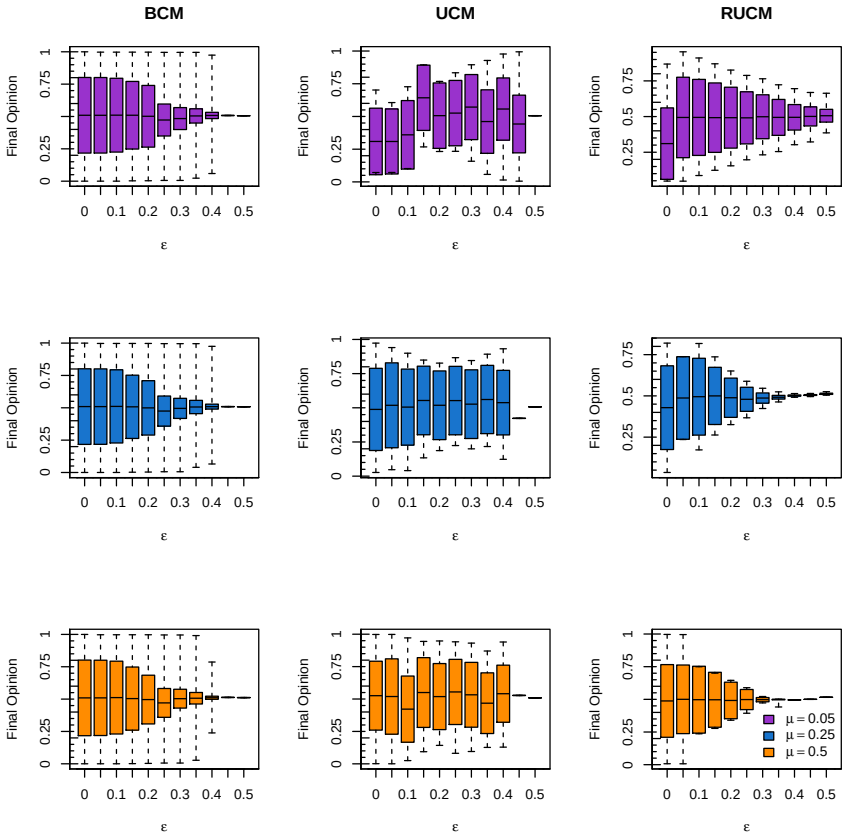


Figure 24: Summary statistics (mean, standard deviation, 1st quantile, and 3rd quantile) of the final opinion distributions for varying ε and three different values of μ: violet denotes $\mu = 0.05$, blue denotes $\mu = 0.25$, and orange denotes $\mu = 0.5$. The left column is for BCM, the central one for UCM, and the right one for RUCM.

5.2.3 Final Distribution of Peaks

We perform Monte Carlo simulations of the BCM, UCM, and RUCM on a scale-free network of 2000 nodes with $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$, for $T = 10^5$ time steps, that are sufficient to reach the final state of the system under the different parameters combinations (the results are averaged over 5 repetitions). Given the final distributions of opinions obtained by the simulations, we compute the number of peaks of opinions as the local maxima in the distribution of frequencies of opinions. To be specific, we divide the interval $[0, 1]$ in 100 bins of length 0.01 and consider the frequencies of values falling in each interval. We regard two peaks to be separate if the distance between the middle points of the respective bins is smaller than 0.1. All the results are averaged over 5 repetitions.

Figure 25 shows the final distribution of peaks for the BCM for varying $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$. The corresponding result for the RBCM model is shown in Fig. 29. The final peaks distribution complies with theoretical (41; 42) and simulation (5) results from previous work. Figure 26 shows the final peaks distribution of UCM (panel a) and RUCM (panel b) for varying $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$. For both models we observe a large area of the parameter space for which two final opinions coexist. We register a faster convergence to the consensus state for the RUCM (w.r.t UCM), that is due to the rewiring rule. Also, we observe that for the RUCM there is a direct transition from many opinions to two opinions, as well as from two opinions to consensus, while for the UCM there is an intermediate area where 3 or 4 opinions emerge, respectively shown in yellow and pale orange.

Comparing Figs. 25 and 26, we see that the new models, unlike the BCM, are able to explain the coexistence of two stable final opinions, often observed in reality. Another important difference with respect to the BCM is that the μ parameter assumes an important role in tuning the number of final opinions peaks. The dependence of the number of final peaks on the μ parameter is stronger for the RUCM, where we observe a clear transition from many opinions to exactly two on the diagonal.

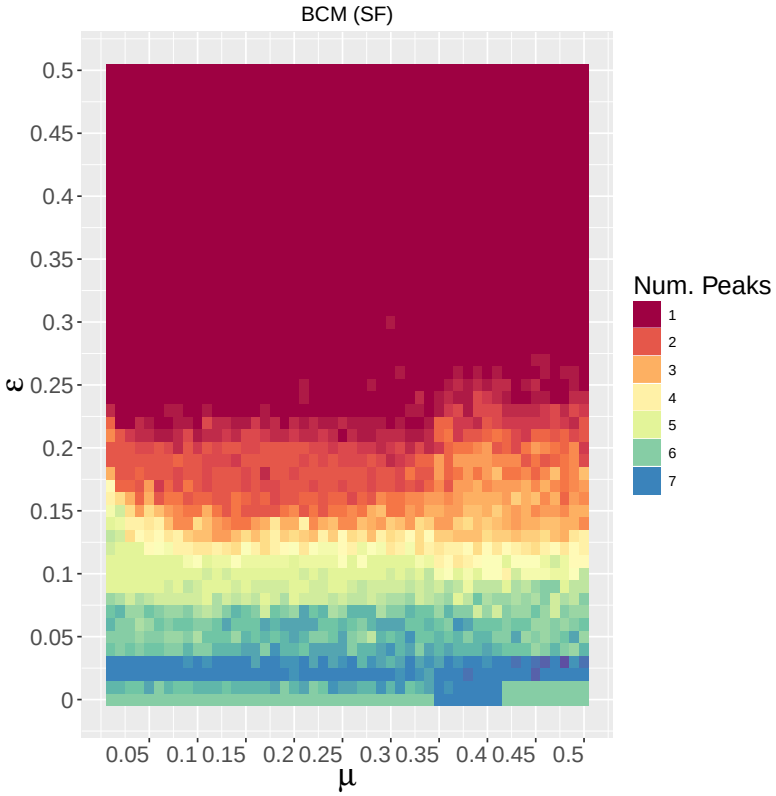
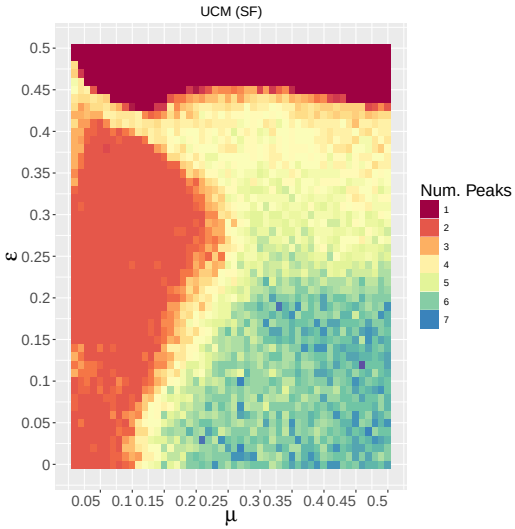
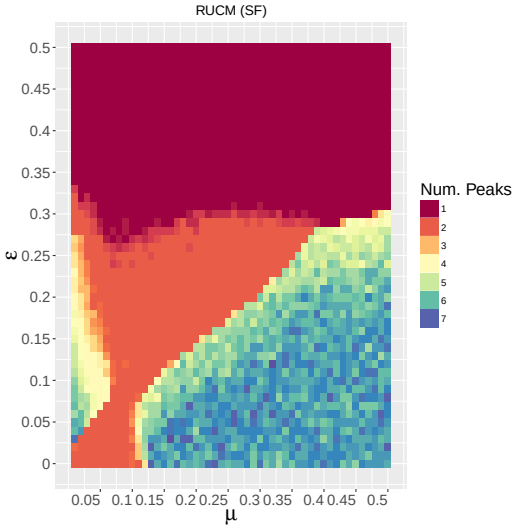


Figure 25: Final peaks distribution for the BCM, with varying $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$. The Monte Carlo simulations are carried on a scale-free network with 2000 nodes for $T = 10^5$ time steps, that are sufficient to reach the final state of the system under the different parameters combinations (all results are averaged over 5 repetitions).



(a)



(b)

Figure 26: Final peaks distributions for the UCM (a) and RUCM (b), with varying $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$. The simulations are carried on a SF network with 2000 nodes for $T = 10^5$ time steps, that are sufficient to reach the final state of the system under the different parameters combinations.

Figure 27 shows the summarizing statistics for the time steps needed to reach the final state (under the different parameters combinations) by boxplots. Black horizontal lines represent the median of the number of steps needed, and the colored boxes represent the interquartile ranges (i.e., the 25th-75th percentile ranges) and they statistically measure the degree of dispersion and the skewness of each analyzed distribution. Vertical lines (i.e., the whiskers) are lower and upper bounded by the minimum and maximum values of the corresponding distribution, once both outliers and extreme values are removed from the data. Individual points represent the outliers of each analyzed distribution. From the left to the right the boxplots refer ot the BCM, RBCM, and UCM.

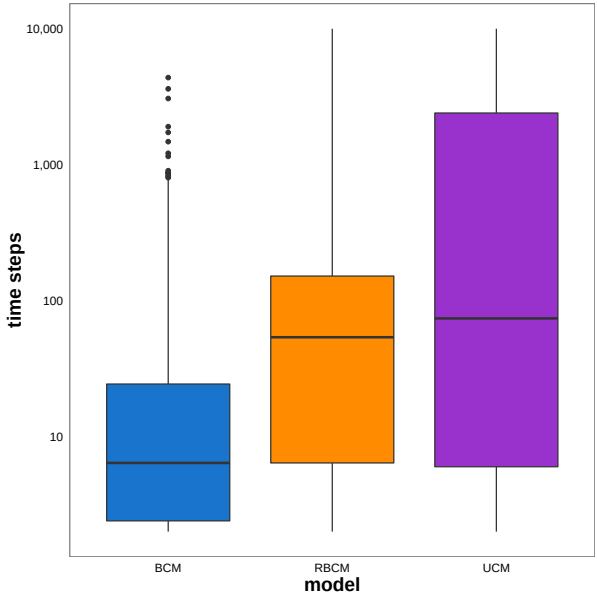


Figure 27: Boxplots of the time steps needed to reach the final state under different combinations of the parameters $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$. From left to right, results for BCM, RBCM, and UCM.

5.2.4 Mean Field Approximation

For the RBCM, after the rewiring steps, all connected agents have an opinion distance below ε, meaning that they will always interact. The time rate of change of $\mathbb{P}(x,t)$ is equal to:

$$\begin{aligned}\frac{\partial \mathbb{P}(x,t)}{\partial t} = & - \mathbb{P}(x,t)\int_{-1}^{1}\mathbb{P}(x+y,t)dy \\ & + \frac{1}{(1-\mu)}\int_{-1-2x}^{1-2x}\mathbb{P}(x+y,t)\mathbb{P}(x-\frac{\mu}{1-\mu}y,t)dy.\end{aligned}$$

Considerations analogous to the BCM case hold (see the Section *Material and Methods*). A faster convergence scale is also observed in the simulations.

In the UCM and RUCM case we consider two updating rules: rule (5.2) if the opinions (x_i, x_j) of the agents are close enough ($|x_i - x_j|_\tau < \varepsilon$) and rule (5.3) if they are not ($|x_i - x_j|_\tau \geq \varepsilon$). Thus the opinions will change according to $(x_i, x_j) \to (\hat{x}_i, \hat{x}_j)$:

$$\begin{aligned}\begin{pmatrix}\hat{x}_i \\ \hat{x}_j\end{pmatrix} = & \begin{pmatrix}1-\vartheta_\varepsilon\mu+(1-\vartheta_\varepsilon)\mu & \vartheta_\varepsilon\mu-(1-\vartheta_\varepsilon)\mu \\ \vartheta_\varepsilon\mu-(1-\vartheta_\varepsilon)\mu & 1-\vartheta_\varepsilon\mu+(1-\vartheta_\varepsilon)\mu\end{pmatrix}\begin{pmatrix}x_i \\ x_j\end{pmatrix} \\ & + (1-\vartheta_\varepsilon)\mu\begin{pmatrix}\varrho(x_j-x_i) \\ \varrho(x_i-x_j)\end{pmatrix},\end{aligned}$$

where $\vartheta_\varepsilon = \vartheta(\varepsilon - |x_i - x_j|_\tau))$ is the Heaviside theta function that equals 1 if $\varepsilon - |x_i - x_j|_\tau < 0$, 0 otherwise, and ϱ is defined in Eq. 5.1). There are two ways in which the density of opinion x changes at every time step t: either an agent moves away from x after an interaction (I^-) or she arrives in x after an interaction (I^+). Let $\mathbb{P}(x,t)dx$ be the fraction of agents whose opinion at time t lies in the interval $[x, x+dx]$, then its time rate of change is:

$$\frac{\partial \mathbb{P}(x,t)}{\partial t} = I^-(x,t) + I^+(x,t).$$

The negative part is defined as in the BCM case but for a wider interval:

$$I^-(x,t) = -\mathbb{P}(x,t)\int_{-1}^{1}\mathbb{P}(x+y,t)dy,$$

as $I^-(x,t)$ is simply the probability that an agent with opinion x interacts with some other agent and thus moves away from x. For $I^+(x,t)$ we have two terms depending on the distance of the initial opinions:

$$I^+(x,t) = I_1^+(x,t) + I_2^+(x,t),$$

for the first term we get the same expression as in the BCM case:

$$I_1^+(x,t) = \frac{1}{(1-\mu)}\int_{-\varepsilon-2x}^{\varepsilon-2x}\mathbb{P}(x+y,t)\mathbb{P}(x-\frac{\mu}{1-\mu}y,t)dy.$$

For $I_2^+(x,t)$ we have to consider the negative update in Eq. (5.3), and the integrals are over the interval for which $|x_1-x_2|_\tau \geq \varepsilon$:

$$\begin{aligned}
I_2^+(x,t) &= \int\int \mathbb{P}(x_1,t)\mathbb{P}(x_2,t)\delta(x+\mu x_1-(1+\mu)x_2-\mu\varrho))dx_1dx_2 \\
&= \frac{1}{(1+\mu)}\int dx_1\mathbb{P}(x_1,t)\int\mathbb{P}(x_2,t)\delta(x_2-\frac{x+\mu x_1-\mu\varrho}{1+\mu})dx_2 \\
&= \frac{1}{(1+\mu)}\int_{|x_1-x_2|_\tau\geq\varepsilon}\mathbb{P}(x_1,t)\mathbb{P}\left(\frac{x+\mu(y-\varrho)}{1+\mu}\right)dx_1 \\
&= \frac{1}{(1+\mu)}\int_{[-1,-\varepsilon-2x]\cup[\varepsilon-2x,1]}\mathbb{P}(x+y,t)\mathbb{P}\left(x+\frac{\mu}{1+\mu}(y-\varrho)\right)dy,
\end{aligned}$$

where $\varrho = \varrho_{x_2-x_1}$. Hence we obtain:

$$\begin{aligned}
\frac{\partial\mathbb{P}(x,t)}{\partial t} = &- \mathbb{P}(x,t)\int_{-1}^{1}\mathbb{P}(x+y,t)dy \\
&+ \frac{1}{(1-\mu)}\int_{-\varepsilon-2x}^{\varepsilon-2x}\mathbb{P}(x+y,t)\mathbb{P}(x-\frac{\mu}{1-\mu}y,t)dy \\
&+ \frac{1}{(1+\mu)}\int_{[-1,-\varepsilon-2x]\cup[\varepsilon-2x,1]}\mathbb{P}(x+y,t)\mathbb{P}\left(x+\frac{\mu}{1+\mu}(y-\varrho)\right)dy.
\end{aligned}$$

When all agents interact positively, i.e. when $\varepsilon \geq 1/2$, the third term of the rate equation disappears and we are again in the BCM case, where

consensus is reached asymptotically and:

$$\mathbb{P}_\infty(x) = M_0\delta(x).$$

From simulations results, we notice that the final state is a single peak as long as $\varepsilon \in (0.45, 0.5)$ for the UCM, or $\varepsilon \in (0.3, 0.5)$ for the RUCM (with the exception of those points for which μ is near to zero).

Unlike the BCM, in the new models the parameter μ plays an important role in the evolution of the distribution of opinions. For both UCM and RUCM we have the coexistence of two opinions in the final state for a wide region of the (ε, μ)-plane, this region varies for the two models, in particular the faster convergence to the consensus state for the RUCM is due to the rewiring rule. For smaller values of ε, and outside the two opinions region, we show by numerical simulations that consensus is not reached, and many opinions at distance larger than ε coexist.

5.2.5 Simulation Results for RBCM

The RBCM differs from the standard BCM in the first phase, where a series of random rewiring steps is performed until all couples in the network agree meaning that the difference between the opinions of the two endpoints of each link is smaller than ε. Through this procedure we obtain a network in which all randomly chosen pairs of users interact and hence the consensus is reached also for smaller values of ε. In Fig. 28 we show the estimated mean number of steps needed to get the fully concordant network as a function of ε, where the results are averaged over 50 realizations. The decay of the estimated mean number of steps is best fitted by the power law ax^{-b}, where the parameters $a = 5.105$ and $b = 1.072$ are obtained through Nonlinear Least Square (NLS) fitting.

Figure 29 shows the final distribution of peaks for the RBCM for varying $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$. We notice that consensus is reached for smaller values of ε w.r.t the BCM. Indeed, while for BCM consensus is reached for $\varepsilon \geq 0.25$, for RBCM we get it for $\varepsilon \geq 0.15$.

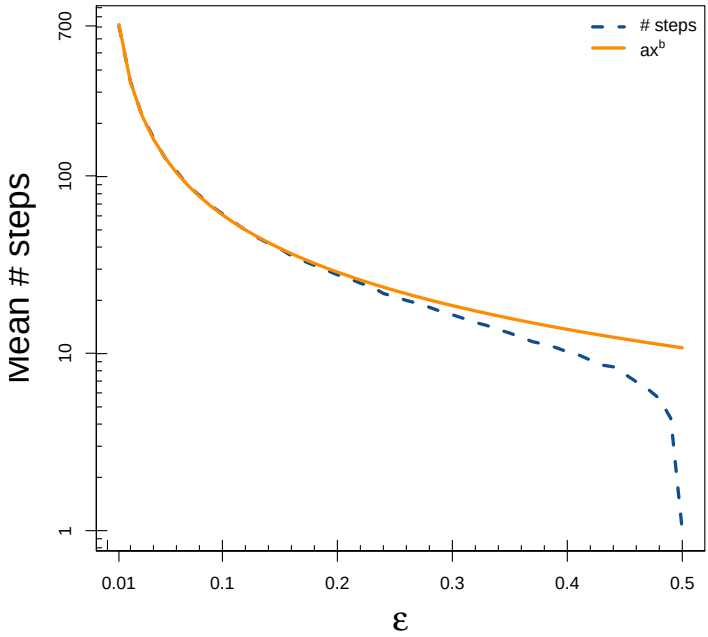


Figure 28: Estimated mean number of steps needed to get the fully agreeing network as a function of ε (dashed blue curve). The results are averaged over 50 realizations. The decay of the estimated mean number of steps is best fitted by the power law ax^{-b} (solid orange curve), where the parameters $a = 5.105$ and $b = 1.072$ are obtained through (NLS) fitting.

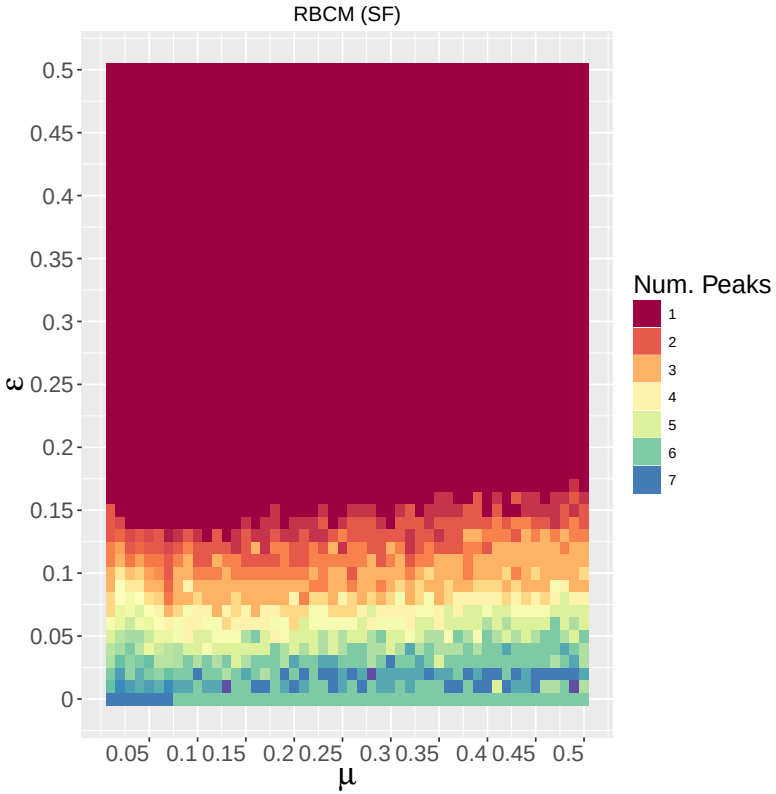


Figure 29: Final peaks distribution for the RBCM, with varying $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$. The Monte Carlo simulations are carried on a scale-free network with 2000 nodes for $T = 10^5$ time steps, that are sufficient to reach the final state of the system under the different parameters combinations (all results are averaged over 5 repetitions).

5.2.6 Simulation Results for ER and SW

In this section we report the simulation results for the Erdös-Rényi random network (ER) and the small-word network (SW). We perform Monte Carlo simulations of the BCM, RBCM, UCM, and RUCM on both ER and SW networks, considering $N = 2000$ nodes and the parameters

$(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$ (the results are averaged over 5 repetitions). Given the final distributions of opinions obtained by the simulations, we compute the number of peaks of opinions as the local maxima in the distribution of frequencies of opinions. To be specific, we divide the interval $[0, 1]$ in 100 bins of length 0.01 and consider the frequencies of values falling in each interval. We regard two peaks to be separate if the distance between the middle points of the respective bins is smaller than 0.1. All the results are averaged over 5 repetitions.

Figure 30 shows the final distribution of peaks of BCM, respectively for ER (a) and SW (b), and RBCM, respectively for ER (c) and SW (d). While, Fig. 31 shows the final peaks distribution of UCM, respectively for ER (a) and SW (b), and RUCM, respectively for ER (c) and SW (d). In all cases we observe a behavior that is qualitatively similar to the SF case.

For both networks RBCM ensures a faster convergence wrt the BCM. Also, UCM and RUCM present a wide area of the parameters space in which two separate opinions coexist. For the newly introduced models we observe a similar behavior on the two different networks and on the scale-free one, in Fig. 26.

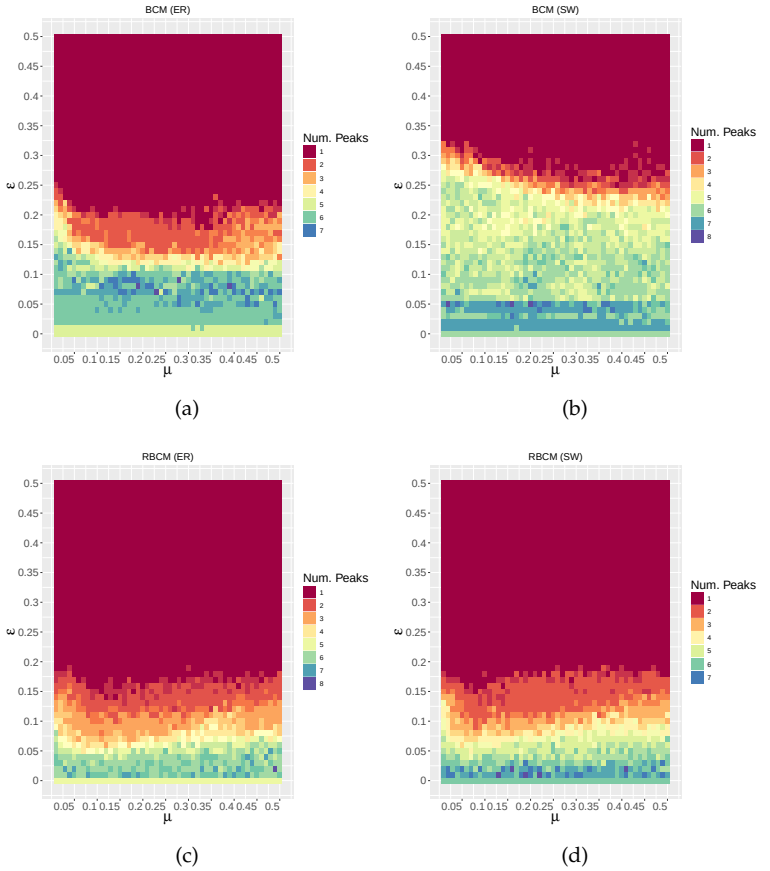


Figure 30: Final distribution of peaks for the BCM, respectively for ER (a) and SW (b), and RBCM, respectively for ER (c) and SW (d), with varying $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$. The Monte Carlo simulations are carried on a Scale-Free network with 2000 nodes for $T = 10^5$ time steps, that are sufficient to reach the final state of the system under the different parameters combinations (all results are averaged over 5 repetitions).

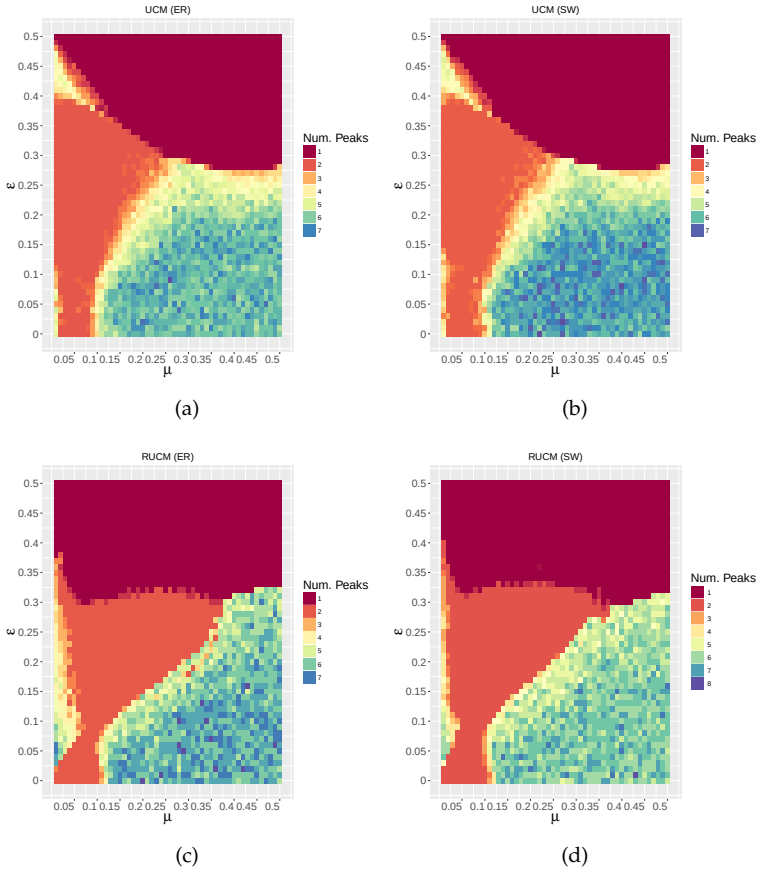


Figure 31: Final distribution of peaks for the UCM, respectively for ER (a) and SW (b), and RUCM, respectively for ER (c) and SW (d), with varying $(\varepsilon, \mu) \in [0, 0.5] \times [0, 0.5]$. The Monte Carlo simulations are carried on a Scale-Free network with 2000 nodes for $T = 10^5$ time steps, that are sufficient to reach the final state of the system under the different parameters combinations (all results are averaged over 5 repetitions).

5.3 Concluding Remarks

In recent years opinion dynamics has attracted much interest from the fields of both statistical physics and social science. In classical models such as the Sznajd model, the voter model, the majority rule model, and the bounded confidence model, consensus is eventually reached,for values of the tolerance parameter big enough. However, in face-to-face and online opinion exchanges, consensus is not commonly achieved, and classical models fail to explain this empirically observed fact.

We propose a model of opinion dynamics capable of reproducing the empirically observed coexistence of two stable opinions. We assume the basic updating rule of the BCM and we develop two variations of the model: the *Rewire with Bounded Confidence Model* (RBCM), in which discordant links are broken until convergence is reached; and the *Unbounded Confidence Model*, under which the interaction among discordant pairs of users is allowed and a negative updating rule is introduced, either with the rewiring step (RUCM) or without it (UCM).

From numerical simulations we find that the new models (UCM and RUCM), unlike the BCM, are able to explain the coexistence of two stable final opinions, often observed in reality. Another important difference with respect to the BCM is that the convergence parameter μ assumes an important role in tuning the number of final opinions peaks; hence, in our model the speed at which opinions converge/diverge allows to change the final opinion landscape. Lastly, we derive a mean field approximation of all the three new models.

Chapter 6

Conclusions and Future Works

In this chapter we outline the main contributions of the thesis and we sketch some possible future works in the same direction.

The works collected in this thesis address different aspects of the online social dynamics, from the spreading of misinformation to the emergence of echo chambers and group polarization. We provide evidences that the diffusion of information, either substantiated or not, is promoted by confirmation bias and homophily. This process in turn generates and fosters the formation of homogeneous polarized clusters, the echo chambers (29). Users' emotional behavior seems to be affected by their engagement within the community. An higher involvement in the echo chamber resolves in a more negative emotional state (29). Putting all results together, we are able to well characterize online information diffusion and the dynamics of polarized groups, inside which users are only reached by information confirming their previous beliefs while they ignore dissenting ones (64). We validate our observations by developing a model of opinions formation that takes into account both confirmation bias and social influence as triggering factors for the group polarization on social networks (31). Our model is able to reproduce the dynamics we observe on Facebook.

However many aspects of opinion formation remain unexplored. Among these we are interested in the dynamics of *non consensus* under which many (more often two) conflicting opinions coexist. In particular, we want to investigate under which conditions minority opinions survive. The problem has been widely explored, a non consensus opinion model (NCO) proposed in (8; 44) allows for the stable coexistence of two opinions by also considering the opinion of the user herself when applying the majority rule update. The model was further refined by investigating the competition between two groups holding dissenting opinions. To this end, authors introduce a set contrarians in one of the two and observe the survival of minorities under this new condition. The survival of a two-opinions state is studied in (14) from a different point of view, considering the emergence of spontaneous recovery of failed nodes and the majority rule update. Both these models assume only two opinion states (± 1) and a majority rule update, with the novelty of accounting for the individual opinion and for an external source of influence.

Conversely, we are interested in exploring the non consensus dynamics in a continuous range of opinions, e.g., with opinions uniformly distributed in the interval $[-1, 1]$, where the opinion indicates a degree of acceptance of a certain topic. We would allow agents to change their opinions under the conjugate influence of the social contacts and some form of mainstream advertising of the topic. Moreover, the opinion update would be either positive, if the distance between two agents' opinions is below a certain threshold, or negative if it is not. The relevance and novelty of this approach lies mainly in the fact that agents may update their opinion of the topic on the basis of both internal (their current opinion) and external (friends, advertising) forces. The dual nature of external forces also embodies two different kinds of influence: social influence and persuasion.

In addition, we are interested in developing new metrics able to capture the criticality of online information, both in terms of their potential virality and their degree of trustworthiness. In the specific, one way to proceed is by characterizing the user response to different information types, and under different environments, i.e., echo chambers. Along this

path, we are interested in exploring the users' response to a wider range of information, that includes informations coming from accredited news sources. A particular emphasis has to be devoted to those topics that can be considered as controversial, e.g., the danger of vaccination, the Brexit.

Appendix A

Other Works

In this section we will report a summary of the main findings produced by our research activity (21; 64; 83; 85; 105; 106; 107), other than those presented in Chapters 3, 4, and 5, at the time of the thesis submission.

The free accessibility of online information, and the possibility for everyone to produce and diffuse personal interpretations of it, is a distinctive feature of our age. Moreover, in the majority of cases online alternative information channels offer unsubstantiated and unverified news. In this hyperconnected and multiple information source environment, getting lost and eventually consume unsubstantiated news is extremely easy. For instance, the World Economic Forum listed *massive digital misinformation* as one of the main risks for the modern society (27; 28).

Recent works (20; 65) analyzed the users' approach toward unsubstantiated claims, underlying the differences between users' systems of believes. Individuals that are prone to consume unverified informations are also more likely to consume in the same way intentionally injected false claims. Our research originates by these previous work and enhances the knowledge about the topic under analysis in terms of finding the driving forces behind viral processes as well as their measurable socio-cognitive determinants, e.g., attention, polarization toward a given content, and testing users' response to debunking campaigns.

A.1 Social Determinants of Content Selection in the Age of (Mis) Information

In this work[1] we perform a quantitative analysis of the information consumption patterns relative to different contents: conspiracy-like theories and scientific news. Our dataset includes about 1.2 millions users on the Italian Facebook and the data collection covers a time period of five years (2010-2014). Conspiracy-like theories tend to reduce the complexity of reality by explaining significant social or political aspects as plots conceived by powerful individuals or organizations. On the other hand science news promote the diffusion of scientific advances, their sources are usually referenced and the contents peer-reviewed. Our analysis targets the quantitative understanding of the social determinants related to content selection, information consumption, and beliefs formation and revision.

We first show the existence of similar consumption patterns of information supporting different (and opposite) worldviews. Then, we measure the social response of polarized users of alternative news to 4502 debunking memes (information aiming at correcting the diffusion of unsubstantiated claims) for increasing level of user engagement, i.e., the number of likes of a user in the category which she belongs to, on the preferred category of information (scientific news and conspiracy-like theories).

Figure 32 shows the quantile discretization of the survival probability distribution for increasing level of users engagement of usual consumers of conspiracy-like theories exposed (panel a) and not exposed (panel b) to debunking memes. We find that polarized users[2] of conspiracy-like claims interacting with debunking are more likely to interact again with conspiracy rumors.

[1] The content in this section is part of the paper (21), published in the proceeding of the International Conference on Social Informatics, 2014. It is a joint work with Alessandro Bessi, Prof. Guido Caldarelli, Dr. Antonio Scala, and Dr. Walter Quattrociocchi.

[2] We say a user to be polarized on science (respectively, on conspiracy) if she left more than 95% of her likes on science (respectively, on conspiracy) posts.

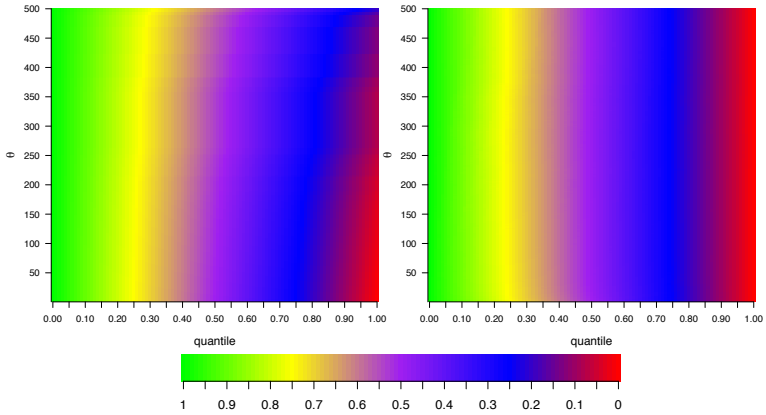


Figure 32: Quantile discretization of the survival probability distribution of conspiracy users against their level of engagement ϑ exposed (panel a) and not exposed (panel b) to posts debunking conspiracy theses.

A.2 Homophily and Polarization in the Age of Misinformation

In this work[3] we analyze a sample of 1.2M Facebook Italian users consuming scientific news and conspiracy-like theories. Our findings show that users' engagement on a specific content correlates with the number of friends having similar consumption patterns (homophily). We then test the relationship between the usual exposure (polarization) to undocumented rumors (conspiracy stories) with respect to the permeability to deliberate false information (4,709 intentional satirical false claims).

Our work provides important insights about the understanding of the diffusion of unverified rumors. Figure 33 shows the log-linear plot of the average fraction $y(u)$ of friends with the same polarization of user u, for users polarized in science (left panel) and conspiracy (right panel). Figure 33 suggests in both cases a linear correlation among the variables;

[3]The content in this section is part of the paper (85), to appear in European Journal of Physics, Special Topics. It is a joint work with Alessandro Bessi, Fabio Petroni, Fabiana Zollo, Dr. Aris Anagnostopoulos, Dr. Antonio Scala, Prof. Guido Caldarelli, and Dr. Walter Quattrociocchi.

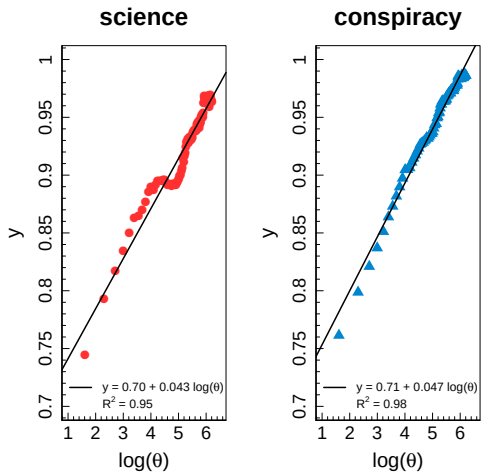


Figure 33: Left panel: scientific polarized users. Right panel: conspiracy polarized users. In both panels, for a polarized user u, we plot the fraction of polarized friends with the same polarization (full red circles for scientific, full blue triangles for conspiracy) with respect to the logarithm of the number of likes $\ln(\vartheta(u))$ of user u. Full lines are the results of a linear regression model $y(u) = \beta_0 + \beta_1 \ln(\vartheta(u))$. Coefficients are estimated using ordinary least squares; in both cases, all the p-values are close to zero.

thus, we check whether for a polarized user u, the fraction of polarized friends in its category $y(u)$ can be predicted by means of a linear regression model where the explanatory variable is a logarithmic transformation of the number of likes $\vartheta(u)$, i.e.

$$y(u) = \beta_0 + \beta_1 \ln(\vartheta(u)).$$

Coefficients are estimated using ordinary least squares, all the p-values are close to zero.

We show that through polarization, we can detect homophily clusters where misleading rumors are more likely to spread. Conversely, such social patterns might represent serious warnings about the effects of current algorithmic specifications for content provisioning. Indeed, the tendency to aggregate in homophile clusters might foster social rein-

forcement and confirmation bias, and thus polarization.

A.3 Modeling Networks with a Growing Feature-Structure

We present a new network model[4] accounting for *homophily* and *triadic closure* in the evolution of social networks. In particular, in our model, each node is characterized by a number of features and the probability of a link between two nodes depends on common features. We do not fix *a priori* the total number of possible features (avoiding a model selection problem for the dimension of the feature-space). The bipartite network of the actors and features evolves according to a stochastic dynamics that depends on three parameters that respectively regulate the preferential attachment in the transmission of the features to the nodes, the number of new features per node, and the power-law behavior of the total number of observed features. We provide theoretical results and statistical estimators for the parameters of the model. We validate our approach by means of simulations and an empirical analysis of a network of scientific collaborations.

A.4 Trend of Narratives in the Age of Misinformation

In this work[5] we analyze a collection of conspiracy-like theories sources in the Italian Facebook over a time span of 4 years. Through a semiautomatic topic extraction strategy, we find that the most discussed contents refer to four well specified semantic categories (or topics): environment, diet, health, and geopolitics. Contents belonging to the different cate-

[4]The content in this section is part of the paper (105), available as pre-print at arXiv:1504.07101. It is a joint work with Prof. Irene Crimaldi, Dr. Greg Morrison, Dr. Walter Quattrociocchi, and Prof. Massimo Riccaboni.

[5]The content in this section is part of the paper (106), published in PloS ONE. It is a joint work with Alessandro Bessi, Fabiana Zollo, Dr. Antonio Scala, Prof. Guido Caldarelli, and Dr. Walter Quattrociocchi.

gories (or topics) are consumed in a very similar way by their respective audience, i.e, users activity in terms of likes and comments on posts belonging to different categories are similar and resolves in comparable information consumption patterns. Conversely, if we focus on the lifetime, i.e., the distance in time between the first and the last comment for each user, we notice a remarkable difference within topics. Users polarized on geopolitics subjects are the most persistent in commenting, whereas the less persistent users are those focused on diet narratives. Finally, by analyzing mobility of users across topics, we find that users can jump independently from one topic to another, and such a probability increases with the user engagement. Users once inside the conspiracy corpus tend to join the overall corpus. This work provides important insights about the fruition of conspiracy-like rumors in online social media and more generally about the mechanisms behind misinformation diffusion.

A.5 Emotional Dynamics in the Age of Misinformation

In this paper[6] we analyze a collection of conspiracy and scientific news sources in the Italian Facebook over a time span of four years (2010 to 2014). We target emotional dynamics inside and across the two polarized communities. In particular, we apply sentiment analysis techniques to the comments of the Facebook posts, and study the aggregated sentiment with respect to scientific and conspiracy-like information. The sentiment analysis is based on a supervised machine learning approach, where we first annotated a substantial sample of comments, and then build a Support Vector Machine classification model.

We focus on the emotional behavior of about 280k Facebook Italian users and perform a quantitative analysis showing that the sentiment on conspiracy pages tends to be more negative than that on science pages. Figure 34 *(top)* shows the proportions of negative, neutral, and posi-

[6]The content in this section is part of the paper (83), published in PloS ONE. It is a joint work with Fabiana Zollo, Dr. Petra Kralj Novak, Alessandro Bessi, Prof. Igor Mozetic, Dr. Antonio Scala, Prof. Guido Caldarelli, and Dr. Walter Quattrociocchi.

tive comments, posts, and users, both on science and conspiracy pages. When considering polarized users we capture an overall increase of the negativity of the sentiment. In Figure 34 *(bottom)* we show the mean sentiment of polarized users as a function of their number of comments. According to our results, the more active polarized users are, the more they tend to be negative, both on science and conspiracy.

A.6 Debunking in a World of Tribes

In this paper[7] we examine the effectiveness of debunking through a quantitative analysis of 54 million users over a time span of five years (Jan 2010, Dec 2014). In particular, we compare how users interact with proven (scientific) and unsubstantiated (conspiracy-like) information on Facebook in the US. As a first step we characterize how distinct types of information – belonging to the two different narratives – get consumed on Facebook. We define the user polarization $\rho \in [-1, 1]$ as the ratio of likes (or comments) on conspiracy posts. In Figure 35 we show that the probability density function (PDF) for the polarization of all the users is a sharply peaked bimodal. The vast majority of users is polarized either towards science ($\rho(u) \sim -1$) or conspiracy ($\rho(u) \sim 1$). Findings confirm the existence of echo chambers where users interact primarily with either conspiracy-like or scientific pages.

To understand whether online debunking campaigns against false rumors are effective, we measure the response of usual consumers of conspiracy stories to 47,713 debunking posts. A first interesting result consists in finding that only a very small fraction of users interacts with debunking posts, i.e., they have commented a debunking post at least once. In Figure 36 we show the survival functions of conspiracy users exposed and not exposed to debunking posts. Notice that the persistence of users[8]

[7]The content in this section is part of the paper (64), available as a pre-print at arXiv:1509.00189. It is a joint work with Fabiana Zollo, Alessandro Bessi, Dr. Antonio Scala, Prof. Guido Caldarelli, Louis Shekhtman, Prof. Shlomo Havlin, and Dr. Walter Quattrociocchi.

[8]The persistence of a user is defined as the number of days between her first and last like (resp., comment) on a conspiracy post.

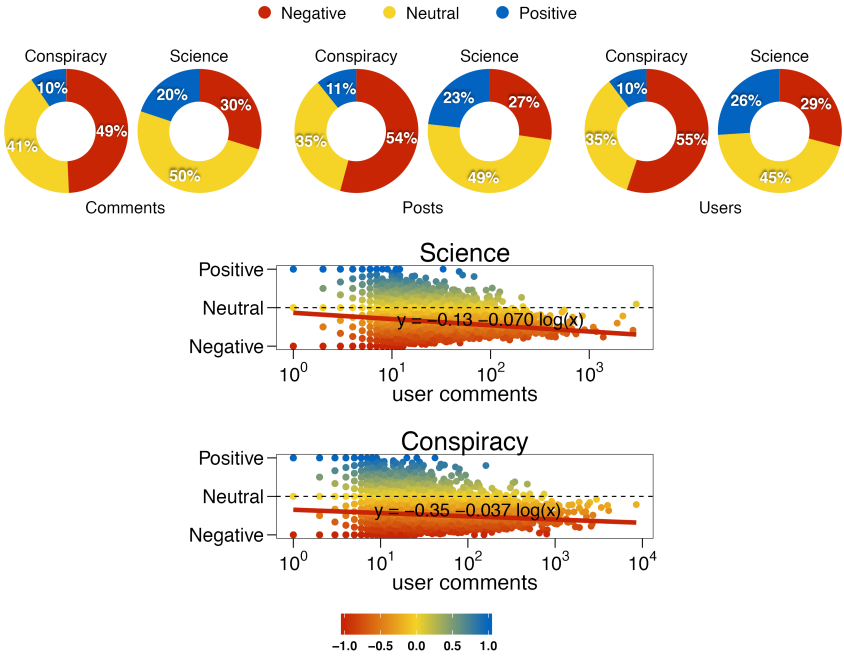


Figure 34: Top: Sentiment on science and conspiracy pages. Proportions of negative, neutral and positive comments (*left*), posts (*center*), and users (*right*) both on science and conspiracy pages. **Bottom:** Sentiment and commenting activity. Average sentiment of polarized users as a function of their number of comments. Negative (respectively, neutral, positive) sentiment is denoted by red (respectively, yellow, blue) color. The sentiment has been regressed w.r.t. the logarithm of the number of comments.

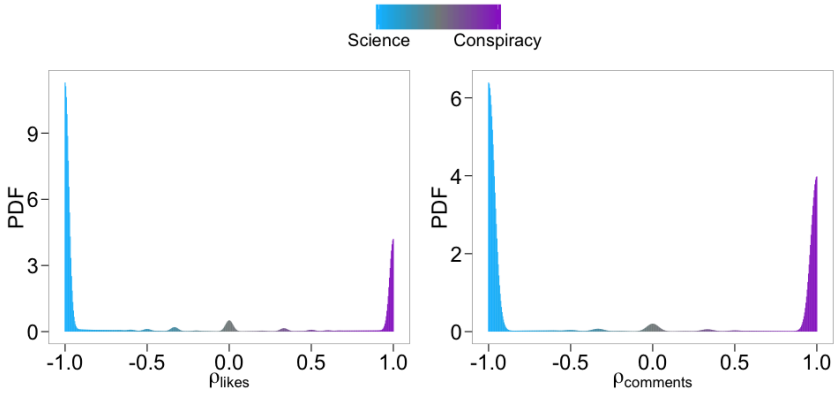


Figure 35: Users polarization. Probability density functions (PDFs) of the polarization of all users computed on likes *(left)* and comments *(right)*.

exposed to debunking is clearly greater than that of not exposed users.

Those few conspiracy users interacting with debunking, rather than internalizing debunking information, often react to it negatively. Indeed, after interacting with debunking posts, users retain, or even increase, their engagement within their echo chamber. According to our results, current debunking campaigns do not seem to be the best options. Our findings suggest that the main problem behind misinformation is conservatism rather than gullibility. When users are faced in online discussion with untrusted opponents the discussion resolves in a major commitment with respect to their own echo chamber.

A.7 User Polarization on Facebook and Youtube

In this paper[9], using a quantitative analysis on a massive dataset (12M of users), we compare consumption patterns of videos supporting scientific and conspiracy-like news on Facebook and Youtube. We extend our analysis by investigating the polarization dynamics, i.e. how users

[9]The content in this section is part of the paper (107), to appear in PloS ONE. It is a joint work with Alessandro Bessi, Fabiana Zollo, Dr. Michelangelo Puliga, Dr. Antonio Scala, Prof. Guido Caldarelli, Dr. Brian Uzzi, and Dr. Walter Quattrociocchi.

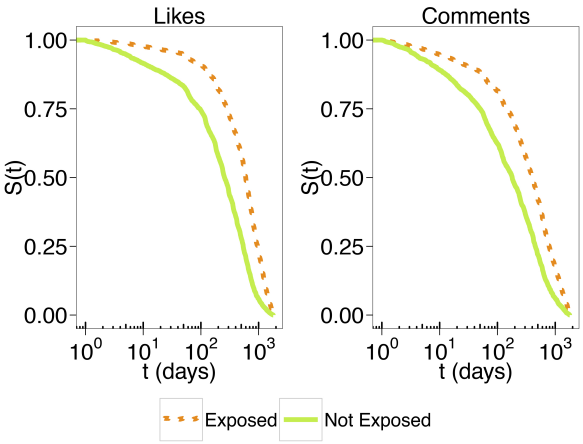


Figure 36: Exposure to debunking. Kaplan-Meier estimates of survival functions of users exposed and not exposed to debunking. Users lifetime is computed both on their likes *(left)* and comments *(right)*.

become polarized comment after comment. On both platforms, we observe that some users interact only with a specific kind of content since the beginning, whereas others start their commenting activity by switching between contents supporting different narratives.

Figure 37 shows the Probability Density Functions (PDFs) of how users distribute their comments on science news and conspiracy-like theories posts (polarization) on both Facebook and YouTube. We observe sharply peaked bimodal distributions. Users concentrate their activity on one of the two narratives. To quantify the degree of polarization we use the Bimodality Coefficient (BC), and we find that the BC is very high for both Facebook and YouTube. Two well separated communities support competing narratives in both online social networks. Content has a polarizing effect, indeed, users focus on specific types of content and aggregate in separated groups independently of the platform and content promotion algorithm.

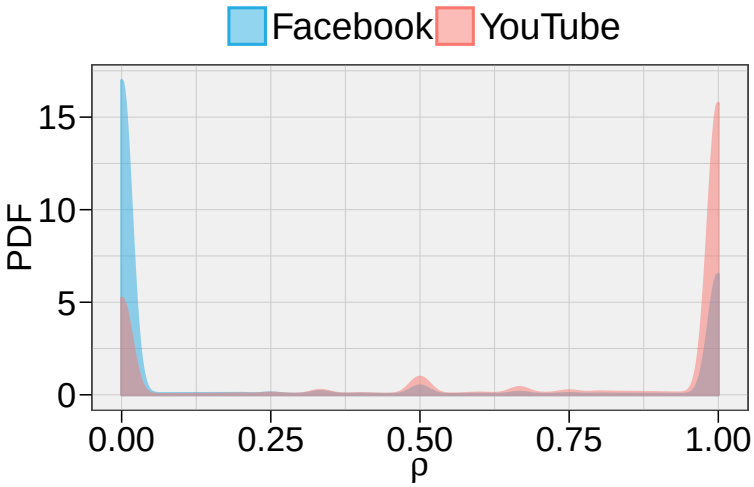


Figure 37: Polarization on Facebook and YouTube. The PDFs of the polarization ϱ show that the vast majority of users is polarized towards one of the two conflicting narratives, i.e. science and conspiracy, on both Facebook and YouTube.

References

[1] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. Computational social science. *Science*, 323(5915):721–723, 2009. 41

[2] Claudio Cioffi-Revilla. Computational social science. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):259–271, 2010.

[3] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591, 2009. 2, 7, 69

[4] Juan Carlos González-Avella, Víctor M Eguíluz, Mario G Cosenza, Konstantin Klemm, JL Herrera, and Maxi San Miguel. Local versus global interactions in nonequilibrium transitions: A model of social dynamics. *Physical Review E*, 73(4):046119, 2006. 2

[5] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04):87–98, 2000. 2, 8, 9, 69, 70, 71, 78

[6] Scott Moss and Bruce Edmonds. Sociology and simulation: Statistical and qualitative cross-validation. *American journal of sociology*, 110(4):1095–1131, 2005. 2

[7] Rosaria Conte, Rainer Hegselmann, and Pietro Terna. *Simulating social phenomena*, volume 456. Springer Science & Business Media, 2013. 2

[8] Jia Shao, Shlomo Havlin, and H Eugene Stanley. Dynamic opinion model and invasion percolation. *Physical review letters*, 103(1):018701, 2009. 2, 10, 69, 92

[9] Dirk Helbing, Illés Farkas, and Tamas Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000. 2

[10] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16):5962–5966, 2012. 2, 10, 11

[11] Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010. 1, 2, 10, 69

[12] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009. 2

[13] Shahar Ronen, Bruno Gonçalves, Kevin Z Hu, Alessandro Vespignani, Steven Pinker, and César A Hidalgo. Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences*, 111(52):E5616–E5622, 2014. 2

[14] Antonio Majdandzic, Boris Podobnik, Sergey V Buldyrev, Dror Y Kenett, Shlomo Havlin, and H Eugene Stanley. Spontaneous recovery in dynamical networks. *Nature Physics*, 10(1):34–38, 2014. 2, 10, 69, 92

[15] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web*, pages 925–936. ACM, 2014. 2, 12

[16] P Alex Dow, Lada A Adamic, and Adrien Friggeri. The anatomy of large Facebook cascades. In *ICWSM*, 2013. 1, 2, 12

[17] Damon Centola and Andrea Baronchelli. The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences*, 112(7):1989–1994, 2015. 1, 69

[18] Maxwell E McCombs and Donald L Shaw. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187, 1972. 1

[19] Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599. Association for Computational Linguistics, 2011. 1

[20] Delia Mocanu, Luca Rossi, Qian Zhang, Márton Karsai, and Walter Quattrociocchi. Collective attention in the age of (mis) information. *Computers in Human Behavior*, 2015. 1, 13, 15, 94

[21] Alessandro Bessi, Guido Caldarelli, Michela Del Vicario, Antonio Scala, and Walter Quattrociocchi. Social determinants of content selection in the age of (mis) information. *International Conference on Social Informatics*, pages 259–268, 2014. 1, 13, 94, 95

[22] Alessandro Bessi, Antonio Scala, Luca Rossi, Qian Zhang, and Walter Quattrociocchi. The economy of attention in the age of (mis) information. *Journal of Trust Management*, 2014. 1, 41, 47

[23] Raymond S Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175, 1998. 2, 3, 69

[24] Leon Festinger. *A theory of cognitive dissonance*, volume 2. Stanford university press, 1962. 2, 3

[25] James O Whittaker. Opinion change as a function of communication-attitude discrepancy. *Psychological Reports*, 13(3):763–772, 1963. 2, 3

[26] Nathaniel Rodriguez, Johan Bollen, and Yong-Yeol Ahn. Collective dynamics of belief evolution under cognitive coherence and social conformity. *arXiv preprint arXiv:1509.01502*, 2015. 3, 10

[27] Lee Howell. Digital wildfires in a hyperconnected world. *WEF Report*, 2013. 3, 15, 94

[28] Walter Quattrociocchi. How does misinformation spread online? In *WEF Agenda*. World Economic Forum, 2016. 3, 15, 41, 94

[29] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016. vii, 5, 14, 41, 46, 68, 69, 70, 73, 91

[30] Michela Del Vicario, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Echo chambers: Emotional contagion and group polarization on facebook. *arXiv preprint arXiv:1607.01032*, 2016. vii, 5, 40, 41, 67

[31] Michela Del Vicario, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. Modeling confirmation bias and polarization. *arXiv preprint arXiv:1509.00189*, 2016. vii, 5, 68, 91

[32] Mark S Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973. 7

[33] Robert Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226, 1997. 7, 8

[34] Richard A Holley and Thomas M Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663, 1975. 7, 69

[35] Thomas M Liggett. Stochastic models of interacting systems. *The Annals of Probability*, 25:1–29, 1997. 7, 69

[36] J Theodore Cox. Coalescing random walks and voter model consensus times on the torus in zd. *The Annals of Probability*, pages 1333–1366, 1989. 7

[37] Serge Galam. Minority opinion spreading in random geometry. *The European Physical Journal B-Condensed Matter and Complex Systems*, 25(4):403–406, 2002. 7

[38] Paul L Krapivsky and Sideny Redner. Dynamics of majority rule in two-state interacting spin system. *Physical Review Letters*, 90, 2003. 7, 69

[39] Katarzyna Sznajd-Weron and Jozef Sznajd. Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(06):1157–1165, 2000. 7, 69

[40] Rainer Hegselmann, Ulrich Krause, et al. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002. 8, 69

[41] Eli Ben-Naim, Paul L Krapivsky, and Sidney Redner. Bifurcations and patterns in compromise processes. *Physica D: Nonlinear Phenomena*, 183(3):190–204, 2003. 9, 72, 73, 78

[42] Eli Ben-Naim and Paul L Krapivsky. Multiscaling in inelastic collisions. *Physical Review E*, 61(1):R5, 2000. 9, 72, 78

[43] Petter Holme and Mark EJ Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E*, 74(5):056108, 2006. 9

[44] Qian Li, Lidia A Braunstein, Shlomo Havlin, and H Eugene Stanley. Strategy of competition between two groups based on an inflexible contrarian opinion model. *Physical Review E*, 84(6):066101, 2011. 10, 69, 92

[45] Armen E Allahverdyan and Aram Galstyan. Opinion dynamics with confirmation bias. *PloS ONE*, 9(7):e99557, 2014. 10

[46] Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007. 10

[47] Elihu Katz and Paul Felix Lazarsfeld. *Personal Influence, The part played by people in the flow of mass communications*. Transaction Publishers, 1966. 11

[48] Robert K Merton. Patterns of influence: Local and cosmopolitan influentials. *Social theory and social structure*, 2:387–420, 1957. 11

[49] Duncan J Watts and Peter Sheridan Dodds. Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458, 2007. 11

[50] Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549, 2009. 11

[51] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528. ACM, 2012. 11

[52] Damon Centola, Juan Carlos Gonzalez-Avella, Victor M Eguiluz, and Maxi San Miguel. Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, 51(6):905–929, 2007. 11

[53] Duncan J Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002. 12

[54] Jure Leskovec, Mary McGlohon, Christos Faloutsos, Natalie Glance, and Matthew Hurst. Cascading behaviour in large blog graphs patterns and a model. *In SDM*, 2007. 12

[55] Yoko Okado and Craig EL Stark. Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory*, 12(1):3–11, 2005. 12

[56] Elizabeth F Loftus. Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4):361–366, 2005. 13

[57] Heike Schmolck, EA Buffalo, and Larry R Squire. Memory distortions develop over time: Recollections of the oj simpson trial verdict after 15 and 32 months. *Psychological Science*, 11(1):39–45, 2000. 13

[58] Steven J Frenda, Rebecca M Nichols, and Elizabeth F Loftus. Current issues and advances in misinformation research. *Current Directions in Psychological Science*, 20(1):20–23, 2011. 13, 39

[59] Aniko Hannak, Drew Margolin, Brian Keegan, and Ingmar Weber. Get back! you don't know me like that: The social mediation of fact checking interventions in twitter conversations. In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014. 13

[60] Eni Mustafaraj Markus Strohmaier Harald Schoen Gayo-Avello, Panagiotis Takis Metaxas, Daniel Peter Gloor, Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Predicting information credibility in time-sensitive social media. *Internet Research*, 23(5):560–588, 2013. 13

[61] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. Correction: Computational fact checking from knowledge networks. *PloS ONE*, 10(10), 2015. 13

[62] Lucas Graves, Brendan Nyhan, and Jason Reifler. Understanding innovations in journalistic practice: A field experiment examining motivations for fact-checking. *Journal of Communication*, 2016. 13

[63] You Wu, Pankaj K Agarwal, Chengkai Li, Jun Yang, and Cong Yu. Toward computational fact-checking. *Proceedings of the VLDB Endowment*, 7(7):589–600, 2014. 13

[64] Fabiana Zollo, Alessandro Bessi, Michela Del Vicario, Antonio Scala, Guido Caldarelli, Louis Shekhtman, Havlin Shlomo, and Walter Quattrociocchi. Debunking in a world of tribes. *arXiv preprint arXiv:1509.00189*, 2015. 13, 41, 68, 91, 94, 100

[65] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Science vs conspiracy: collective narratives in the age of misinformation. *PloS one*, 10(2):02, 2015. 13, 15, 41, 42, 47, 68, 94

[66] Cass Sunstein. *Echo chambers*. Princeton Univ Press, Princeton, NJ, 2001. 15

[67] R Kelly Garrett. Echo chambers online?: Politically motivated selective exposure among internet news users1. *Journal of Computer-Mediated Communication*, 14(2):265–285, 2009. 15

[68] Facebook. Using the graph API. Website, 8 2013. last checked: 19.01.2014. 16, 43

[69] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *Nature*, 393(6684):440–442, 1998. 26, 71, 74

[70] Bi Zhu, Chuansheng Chen, Elizabeth F Loftus, Chongde Lin, Qinghua He, Chunhui Chen, He Li, Robert K Moyzis, Jared Lessard, and Qi Dong. Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual Differences*, 48(8):889–894, 2010. 39

[71] James Surowiecki. The wisdom of crowds: why the many are smarter than the few. london. *Abacus: New Edition*, 2005. 39

[72] R Kelly Garrett and Brian E Weeks. The promise and peril of real-time corrections to political misperceptions. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1047–1058. ACM, 2013. 39

[73] Michelle L Meade and Henry L Roediger. Explorations in the social contagion of memory. *Memory & cognition*, 30(7):995–1009, 2002. 39

[74] Asher Koriat, Morris Goldsmith, and Ainat Pansky. Toward a psychology of memory accuracy. *Annual review of psychology*, 51(1):481–537, 2000. 39

[75] Michael S Ayers and Lynne M Reder. A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, 5(1):1–21, 1998. 39

[76] Michael A Cacciatore, Dietram A Scheufele, and Shanto Iyengar. The end of framing as we know it and the future of media effects. *Mass Communication and Society*, 19(1):7–23, 2016. 41

[77] Jo Brown, Amanda J Broderick, and Nick Lee. Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of interactive marketing*, 21(3):2–20, 2007. 41

[78] Richard Kahn and Douglas Kellner. New media and internet activism: From the'battle of seattle'to blogging. *New media & society*, 6(1):87–95, 2004. 41

[79] Walter Quattrociocchi, Rosaria Conte, and Elena Lodi. Opinions manipulation: Media, power and gossip. *Advances in Complex Systems*, 14(04):567–586, 2011. 41

[80] Walter Quattrociocchi, Guido Caldarelli, and Antonio Scala. Opinion dynamics on interacting networks: media competition and social influence. *Scientific reports*, 4, 2014. 41, 70

[81] Ravi Kumar, Mohammad Mahdian, and Mary McGlohon. Dynamics of conversations. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 553–562. ACM, 2010. 41

[82] Cass R Sunstein. The law of group polarization. *Journal of political philosophy*, 10(2):175–195, 2002. 41, 68

[83] Fabiana Zollo, Petra Kralj Novak, Michela Del Vicario, Alessandro Bessi, Igor Mozetic, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Emotional dynamics in the age of misinformation. *PloS ONE*, 10(9):e0138740, 2015. 41, 46, 56, 68, 94, 99

[84] Alessandro Bessi, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Viral misinformation: The role of homophily and polarization. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 355–356. International World Wide Web Conferences Steering Committee, 2015. 41, 68

[85] Alessandro Bessi, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Homophily and polarization in the age of misinformation. *(to appear on) Eur. Phys. J. S.T.*, 2016. 16, 41, 94, 96

[86] L Ferrante, S Bompadre, L Possati, and L Leone. Parameter estimation in a gompertzian stochastic model for tumor growth. *Biometrics*, 56(4):1076–1081, 2000. 43, 44

[87] Dragan Juki, Gordana Kralik, and Rudolf Scitovski. Least-squares fitting gompertz curve. *Journal of Computational and Applied Mathematics*, 169(2):359 – 375, 2004. 43, 44

[88] Michael Ghil, MR Allen, MD Dettinger, K Ide, D Kondrashov, ME Mann, Andrew W Robertson, A Saunders, Y Tian, F Varadi, et al. Advanced spectral methods for climatic time series. *Reviews of geophysics*, 40(1), 2002. 45, 51

[89] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for time series*. Springer Science & Business Media, 2013. 45, 51

[90] Silvia Maria Alessio. *Digital Signal Processing and Spectral Analysis for Scientists: Concepts and Applications*. Springer, 2016. 45

[91] Myles R Allen and Leonard A Smith. Monte Carlo SSA: Detecting irregular oscillations in the presence of colored noise. *Journal of Climate*, 9(12):3373–3404, 1996. 45, 53

[92] Andreas Groth and Michael Ghil. Monte Carlo singular spectrum analysis (SSA) revisited: detecting oscillator clusters in multivariate datasets. *Journal of Climate*, 28(19):7873–7893, 2015. 45, 53

[93] John M Chambers. *Graphical methods for data analysis*. Wadsworth International Group, 1983. 47

[94] Walter Quattrociocchi, Antonio Scala, and Cass R Sunstein. Echo chambers on Facebook. *Available at SSRN*, 2016. 68

[95] Audun Jøsang, Walter Quattrociocchi, and Dino Karabeg. Taste and trust. In *IFIP International Conference on Trust Management*, pages 312–322. Springer, 2011. 68

[96] Stefan König, Tina Balke, Walter Quattrociocchi, Mario Paolucci, and Torsten Eymann. On the effects of reputation in the internet of services. In *Proceedings of the 1st Int. Conference on Reputation (ICORE 2009)*, pages 200–214, 2009. 69

[97] M Paolucci, T Eymann, W Jager, J Sabater-Mir, R Conte, S Marmo, S Picascia, W Quattrociocchi, T Balke, S Koenig, et al. Social knowledge for e-governance: Theory and technology of reputation. *Roma: ISTC-CNR*, 2009. 69

[98] Renaud Lambiotte and Sidney Redner. Dynamics of non-conservative voters. *EPL (Europhysics Letters)*, 82(1):18007, 2008. 69

[99] Serge Galam. Sociophysics: a review of galam models. *International Journal of Modern Physics C*, 19(03):409–440, 2008. 69

[100] Jan Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838, 2007. 69, 71

[101] Marlon Ramos, Jia Shao, Saulo DS Reis, Celia Anteneodo, José S Andrade, Shlomo Havlin, and Hernán A Makse. How does public opinion become extreme [quest]. *Scientific reports*, 5, 2015. 69

[102] Michele Starnini, Mattia Frasca, and Andrea Baronchelli. Emergence of metapopulations and echo chambers in mobile agents. *arXiv preprint arXiv:1603.04789*, 2016. 70

[103] Paul Erdős and Alfréd Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959. 71, 74

[104] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. 71, 74

[105] Irene Crimaldi, Michela Del Vicario, Greg Morrison, Walter Quattrociocchi, and Massimo Riccaboni. Modeling networks with a growing feature-structure. *arXiv preprint arXiv:1504.07101*, 2015. 94, 98

[106] Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Trend of narratives in the age of misinformation. *PloS ONE*, 10(8):e0134641, 2015. 94, 98

[107] Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi. Users polarization on Facebook and Youtube. *(to appear on) PloS ONE*, 2016. 94, 102