IMT School for Advanced Studies, Lucca

Lucca, Italy

Emotional Contagion and Group Polarization: Experimental Evidence on Facebook

PhD Program in Computer, Decision, and Systems Science

XXIX Cycle

By

Fabiana Zollo

2016

The dissertation of Fabiana Zollo is approved.

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

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

Co-Advisor:
Prof. Guido Caldarelli, IMT School for Advanced Studies Lucca

The dissertation of Fabiana Zollo has been reviewed by:

Dr. Andrea Gabrielli, Institute for Complex Systems, CNR, Italy

Dr. Sabrina Tiziana Gaito, Department of Computer Science, Università degli Studi di Milano, Italy

IMT School for Advanced Studies, Lucca

2016

Prendete la vita con leggerezza,
che leggerezza non è superficialità,
ma planare sulle cose dall'alto,
non avere macigni sul cuore.

Italo Calvino

To Duccio
and all the amazing people
who supported me
in the bad days
of the last three years.

CONTENTS

Acknowledgements

I am sincerely grateful to *Walter Quattrociocchi* for involving me in such an exciting research work, for his absolute trust in me, and for recalling me the importance of not taking life too seriously.

I would like to thank *Sabrina Tiziana Gaito* and *Andrea Gabrielli* for reviewing this thesis and providing interesting insights and comments.

I am also thankful to *Igor Mozetič* for sharing his knowledge and technical know-how and for his constant support during my visiting fellowship at Jožef Stefan Institute of Ljubljana, Slovenia.

Finally, a thank you to *Antonio Scala*, for his enthusiasm, support, and precious suggestions, and to all my co-authors, with a special thought for *Michela Del Vicario*, precious colleague, pleasant roommate, and sincere friend.

Vita

October 24, 1987	Benevento (BN), Italy
2010	B.Sc. in Computer Science Final mark: 101/110 University of Perugia, Perugia, Italy
2013	M.Sc. in Computer Science Final mark: 110/110 cum laude University of Perugia, Perugia, Italy

Publications

1. M. Del Vicario, G. Vivaldo, A. Bessi, **F. Zollo**, A. Scala, G. Caldarelli, and W. Quattrociocchi. "Echo Chambers: Emotional Contagion and Group Polarization on Facebook", preprint available at *arXiv:1607.01032*, (Under Review), 2016
2. A. Bessi, **F. Zollo**, M. Del Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, and W. Quattrociocchi. "Users Polarization on Facebook and Youtube", in *PLoS ONE 11(8)*, 2016.
3. A. Bessi, F. Petroni, M. Del Vicario, **F. Zollo**, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi. "Homophily and Polarization in the Age of Misinformation", (to appear) in *Eur. Phys. J. Special Topics*, 2016.
4. M. Del Vicario, A. Bessi, **F. Zollo**, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, W. Quattrociocchi. "The Spreading of Misinformation Online", in *Proceedings of the National Academy of Sciences (PNAS)*, 2016.
5. **F. Zollo**, A. Bessi, M. Del Vicario, A. Scala, G. Caldarelli, L. Shekhtman, S. Havlin, W. Quattrociocchi. "Debunking in a World of Tribes", preprint available at *arXiv:1510.04267*, (Under Review), 2015
6. **F. Zollo**, P.K. Novak, M. Del Vicario, A. Bessi, I. Mozetic, A. Scala, G. Caldarelli, and W. Quattrociocchi. "Emotional Dynamics in the Age of Misinformation", in *PLoS ONE 10(9)*, 2015
7. A. Bessi, **F. Zollo**, M. Del Vicario, A. Scala, G. Caldarelli, and W. Quattrociocchi. "Trend of Narratives in the Age of Misinformation", in *PLoS ONE 10(8)*, 2015
8. A. Costantini, O. Gervasi, **F. Zollo**, and L. Caprini. "User Interaction and Data Management for Large Scale Grid Applications", in *Journal of Grid Computing*, 2014
9. **F. Zollo**, L. Caprini, O. Gervasi, A. Costantini. "X3DMMS: an X3DOM Tool for Molecular and Material Sciences", in *Proceedings of the 2011 Web3D ACM Conference*, 129-136, 2011

Abstracts & Posters

1. B. Sluban, **F. Zollo**, G. Caldarelli, I. Mozetič, and W. Quattrociocchi. "Social dynamics of online debates on unverified news", at *Conference on Complex Systems*, Amsterdam, The Netherlands, 2016
2. A. Bessi, M. Del Vicario, **F. Zollo**, A. Scala, G. Caldarelli, W. Quattrociocchi. "Misinformation on Online Social Media", at *International Conference on Computational Social Science*, Helsinki, Finland, 2015
3. A. Bessi, F. Petroni, M. Del Vicario, **F. Zollo**, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi. "Viral misinformation: the role of homophily and polarization", in *Proceedings of WWW'15, WebSci Track Papers & Posters*, 2015

Presentations

1. F. Zollo, "Community dynamics in the age of credulity", at *Festival delle Comunità del Cambiamento*, Milano, Italy, October 2016.
2. F. Zollo, "Social dynamics of online debates on unverified news", at *Conference on Complex Systems*, Amsterdam, The Netherlands, September 2016.
3. F. Zollo, "Social media dynamics in the age of credulity," at *International Journalism Festival*, Perugia, Italy, April 2016.
4. F. Zollo, "X3DMMS: an X3DOM Tool for Molecular and Material Sciences," at *Web3D ACM Conference*, Paris, France, June 2011.

Media Coverage, Interviews, & Reports

1. *Perché smentire le bufale è inutile*, La Stampa Tecnologia, Mar 16 2016
2. *False informazioni e complottismo, ecco lo studio che spiega come funzionano*, La Nazione, Feb 23 2016
3. *Il dibattito sui vaccini approda su Pinterest*, OggiScienza, Feb 11 2016

4. *L'era della disinformazione* di Walter Quattrociocchi, Le Science n.570, Feb 2016
5. *How does misinformation spread online?* by Walter Quattrociocchi, World Economic Forum, Jan 14 2016
6. *How Facebook Makes Us Dumber* by Cass R. Sunstein, Bloomberg View, Jan 8 2016
7. *An hour-by-hour look at how a conspiracy theory becomes 'truth' on Facebook* by C. Dewey, The Washington Post, Jan 7 2016
8. *La disinformazione corre sui social media*, Le Scienze, Jan 5 2016
9. *Facebook study suggests online users reinforce their views by creating echo chambers* by Bob Yirka, Phys.org, Jan 5 2016
10. *The truth is rushing out there: why conspiracy theories spread faster than ever*, The Guardian, Dec 26 2015
11. *What was fake on the Internet this week: Why this is the final column* by C. Dewey, The Washington Post, Dec 18 2015
12. *Bufale online e Italia credulona*, La Stampa, Nov 10 2015
13. *Los bulos en Facebook resisten a los esfuerzos por desmontarlos*, El Pais, Nov 2 2015
14. *L'amaca* di Michele Serra, La Repubblica, Oct 31 2015
15. *Smentire le bufale è inutile?*, Il Post, Oct 30 2015
16. *What was fake on the Internet this week: Why do we even bother, honestly?*, The Washington Post, Oct 30 2015
17. *La calunnia per fermare il cambiamento* di Gianni Riotta, La Stampa Opinioni, Oct 22 2015
18. *Continueremo a credere a notizie false, i cacciatori di bufale da soli non bastano*, Repubblica.it Tecnologia, Oct 20 2015
19. *Why it's impossible to debunk a social-media conspiracy theory*, The Daily Dot, Oct 16 2015
20. *Le dinamiche emotive tra scienza e complotto su Facebook*, OggiScienza, Oct 13 2015
21. *Perché cerchiamo di smontare le bufale*, Wired, Oct 2 2015
22. *Disintermediation: Digital Wildfires in the Age of Misinformation* by Alessandro Bessi and Walter Quattrociocchi, Australian Quarterly, Oct-Dec 2015
23. *Why your Facebook news feed is filled with conspiracy theories*, The Daily Dot, Apr 22 2015
24. *Come ti costruisco una bufala sul web*, Repubblica.it Inchieste, Jan 8 2015

Abstract

Information, rumors, debates shape and reinforce the perception of reality and heavily impact public opinion. Indeed, the way in which individuals influence each other is one of the foundational challenges in several disciplines such as sociology, social psychology, and economics. One of the most fascinating and powerful mechanisms of social contagion is that of group polarization. The phenomenon manifests when like-minded people discuss and reinforce their shared views thus ending up in a more extreme position. The core of the research work presented in this thesis explores the phenomenon of group polarization on online social media. We focus on the Italian and US pages providing scientific and conspiracy information and we analyze a) users' emotional dynamics and b) their response to dissenting information. We offer tight quantitative evidence about the existence of echo chambers on online social media. Users tend to promote their beliefs and to form highly polarized groups. Furthermore, dealing with untrusted opponents in online discussion results for users in a major commitment with respect to their own echo chamber.

CHAPTER

ONE

INTRODUCTION

1.1 Background

Information, rumors, debates shape and reinforce the perception of reality and heavily impact public opinion. The way in which individuals influence each other, indeed, is one of the foundational challenges in several disciplines such as sociology, social psychology, and economics. One of the most fascinating and powerful mechanisms of social contagion is that of group polarization. The phenomenon manifests when like-minded people discuss and reinforce their shared views thus ending up in a more extreme position. The core of the research work presented in this thesis explores, by means of quantitative methods, the phenomenon of group polarization on online social media.

Global communications have been extremely facilitated by the rapid advance of the Internet and web technologies; digital interactions allow news and information to spread all over the world rapidly and intensively. These changes have dramatically influenced the way information get consumed, leading up to the formation of a scenario where few suppliers provide fact-checked information (e.g., publishers, news organisations, the academy). In turn, a heterogeneous mass of alternative information sources emerged, fostered by the active participation of people in the production and diffusion of contents.

Such a large diversification of information deeply affects the mechanisms behind the formation of public opinion (1; 2; 3); the active role played by people determined the emergence of new knowledge, enthusiastically dubbed as *collective intelligence* (4; 5; 6; 7). Nevertheless, social media are pervaded by the presence of unsubstantiated or untruthful rumours resulting in a sort of *collective credulity*. As stated in (8), people are misinformed when they confidently hold wrong beliefs. Indeed, *misinformation* might negatively influence the public opinion.

The empirical investigations conducted in (8) show that, in general, people tend to resist facts, holding inaccurate factual beliefs confidently. Indeed, in 2013 WEF placed the global risk of massive digital misinformation at the core of technological and geopolitical risks ranging from terrorism to cyber attacks and the failure of global governance (9). Moreover, results in (10) also indicate that corrections frequently fail to reduce misperceptions; on the contrary, in several cases they act as a *backfire effect*, actually increasing misperceptions among the considered group.

Thus, beyond its great benefits, a hyperconnected world might allow the viral spread (i.e., a rapid and wide diffusion of a piece of information among Internet users) of misleading or provocative information, that could result in serious real-word consequences. Indeed, such a scenario represents a florid environment for digital wildfires – i.e, viral phenomena triggered by false or sensitive information online – when combined with both functional illiteracy and *confirmation bias* – i.e., the tendency to search, select, and interpret information coherently with one's system of beliefs.

For instance, it has been reported that inadequate health policies in South Africa led to more than 300,000 unnecessary AIDS deaths (11); however, these tragic events has been exacerbated by AIDS *denialists*, who state that HIV is inoffensive and that antiretroviral drugs cause, rather than treat, AIDS. In fact, several works pointed out the dangers of denying AIDS (12; 13); as Kalichman asserts, *AIDS denialism is the outright rejection of science and medicine [and] has emerged as a genuine menace to global public health including in the United States and, particularly, in South*

Africa. Similar considerations could be extended to the recent Ebola outbreak in west Africa. After the death of two people having drunk salt water, the World Health Organisation (WHO) restated that all rumours about hypothetical cures or practices are false and that their use can be dangerous (14). More recently, we have witnessed the American case of Jade Helm 15, a military training exercise which took place in multiple US states. The drill turned out to be perceived as a conspiracy plot aiming at imposing martial law, to the extent that the Texas Gov. Greg Abbott ordered the State Guard to monitor the operations.

Consequently, several concerns have been expressed about social influence on the Internet. As an example, Donald Trump is the Republican Party nominee in the 2016 US presidential election, against all odds. His strategic social media campaign is still a matter of speculation, and many hypothesis have been put forward about the reasons behind its success and the impact on the electoral process. Similar doubts have been raised during the Brexit – the British referendum to leave the European Union – campaign, where both sides, *Leave* and *Remain*, battled it out on social media. Indeed, on the Internet people can access more and more extreme versions of their own judgements. In this way the benefits coming from exposure to different points of view can be dramatically reduced (15). Individuals, and the groups that they form, may move to a more extreme point in the same direction indicated by their own preexisting beliefs; when people discuss with many like-minded others, their views become more extreme (16). This phenomenon is known as *group polarization* and is directly connected to concerns about the consequences of the Internet, which allows for a relative anonymity and may intensify group polarization (17). Indeed, social influence is one of the main mechanisms underlying group polarization and may affect an individual's behavior in two different ways, 1) informational and 2) reputational (18; 19). In the first case, what other (relevant) people do or say carries an *informational externality* (20) and observers are willing to follow them. In the second case, observers may do what they believe other people think they should do, just because they care about their reputation.

A very famous experiment about group influences was conducted by

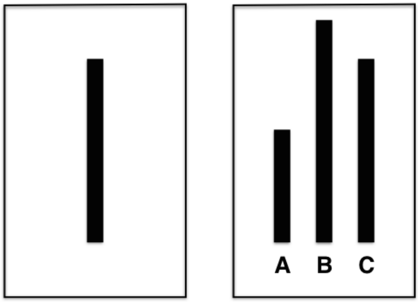


Figure 1: A replication of the cards used during the Asch experiment. The card on the left is for reference, the one on the right shows the comparison lines.

Solomon Asch in 1955 (21). The task of the subjects was very simple: they had to match a certain line placed on a white card with the corresponding one – i.e., having the same length – among three other lines placed on another white card. A replication of the cards is shown in Fig. 1. The subject was one of the eight people taking part to the test, but was unaware that the others were there as part of the research. The experiment consisted of three different rounds. In the first two rounds everyone provided the right (and quite obvious) answer. In the third round some group members matched the reference line to the shorter or longer one on the second card, introducing the so-called *unexpected disturbance* (22). In this case the subject could decide to keep his decision unchanged or yield to the others. Normally subjects erred less than 1% of the time; but in these rounds they erred 36.8% of the time (23). The experiment showed that, under group pressure, individuals were highly likely to abandon the direct evidence of their own senses. Both informational and reputational considerations appear to have led people toward these errors (16).

Another relevant study was conducted by James Stoner, who identified the so-called *risky shift* (24). In the experiment people were first asked to study twelve different problems and provide their personal judgement; after that, they had to join as a group and take a final decision together. Out of thirteen groups, twelve repeatedly showed a pattern towards greater risk-taking. While 45% did not change their judgements at all, only 16% were moved in the direction of greater caution and 39% in that of greater risk-taking. This shift is the above-named risky shift.

Hence, group discussion may enhance the initial tendency of individual group members and fosters polarization. In general, two main explanations may be listed for group polarization: social comparison and persuasive arguments. In the first case, people want to be perceived favorably both by other group members and by themselves; consequently, they adjust their judgement towards the dominant position. In the second case, the individual's choice shifts in the direction of the most persuasive position presented within the group. Therefore, when a group is formed by members who are already inclined in a certain direction, the majority of the arguments supports the same direction. As a result, the decision is likely to move individuals further in the direction of their initial judgements.

1.2 Advances

This thesis aims at studying quantitatively the phenomenon of group polarization on online social media. The emerging field of computational social science (CSS) (25; 26) benefits from the large availability of data from online social networks. Indeed, social scientists were limited until now to surveys (that are backdated) and lab experiments, that are usually performed on a relatively small group of people and deeply affected by external validity issues. In (27) authors argued that both sample surveys and in-depth interviews are dated research methods by now. CSS, instead, leverages the capacity to collect and analyze data at unprecedented scales and levels of details and may allow to discover patterns of individual and group behaviours.

However, a series of challenges has to be considered, as pointed out by Marc Huberty in (28). First, we can not claim that our data allow a clear and unbiased study of humanity; most of the data comes from services such as web pages, online shopping or social media, the so-called *digital exhaust* (29), which does not concern the society in general. Therefore, the uses of such data are limited. Second, we can not assert that understanding online behaviour today implies that we will be able to do the same tomorrow. Third, even if online and offline identity are not completely separate, researchers have shown that individuals' online identities vary from their offline selves largely; hence, we can not state that online behaviour coincides with offline behaviour. Finally, we can not assume that complex patterns of social behaviour today will still describe the world that we would like to predict tomorrow.

Nevertheless, the development of tools for understanding and analyzing social dynamics on the web takes on great importance and involves a cross-methodological approach to formulate and validate data-driven models. Recent works such as (30; 31; 32; 33) focused on structural properties of the network to determine the way in which news spread in social networks, what makes messages go viral and what are the characteristics of users who help spread such information. `Snopes.com` (34), launched as a purely urban-legends site in 1995, is considered one of the most reliable resources for sifting through political and media facts and fallacies, especially since the Sept. 11 terrorist attacks. In 2014 journalist Craig Silverman also launched `Emergent.info` (35), a website aiming at containing the viral spread of unsubstantiated news by monitoring web contents in real time. Nonetheless, both of them do not take into consideration two significant factors: first, people who are not interested in knowing the truth about a certain topic will probably consider its denial as a further attempt to control information; second, the system does not consider the complexity of some information and the essential role played by socio-cognitive factors in this kind of mechanisms.

Here we present a collection of works that address the phenomenon of group polarization on Facebook. Focusing on the Italian and US pages providing scientific and conspiracy information we analyse a) users' emotional dynamics and b) their response to dissenting information. The main results of the thesis may be summarized as follows:

1. By means of sentiment analysis techniques, we show that users committed to the same narrative tend to negatively influence each

other. Such an emotional influence is even more negative when opposite factions of users meet and discuss.

2. The tendency of users to join polarized groups sharing the same narrative creates a segregation effect and dissenting information is mainly ignored.

We offer tight quantitative evidence about the existence of echo chambers on online social media. Users tend to promote their beliefs and form highly polarized groups. Furthermore, dealing with untrusted opponents in online discussion results for users in a major commitment with respect to their own echo chamber. As an example, by examining the response of users to debunking efforts, we find that these attempts are largely ineffective and only serve to reinforce people's preexisting beliefs (36).

The thesis is structured as follows: in Chapter 2 an overview of the state-of-the-art is provided, especially focusing on information diffusion, emotional contagion, and rumors spreading; in Chapters 3 and 4 main research works are discussed; in Chapter 5 conclusions are drawn and future works are sketched; finally, in Appendix A secondary works are briefly presented, while details about Facebook datasets are given in Appendix B.

CHAPTER
TWO

STATE-OF-THE-ART

2.1 Social, Emotional, and Informational Contagion

In 2009 a paper on Science (26) proclaims the birth of the Computational Social Science (CSS), an emerging research field aiming at studying massive social phenomena quantitatively by means of a multidisciplinary approach based on Computer Science, Statistics, and Social Sciences. In (25) the author stresses the importance to conduct studies aiming at understanding the structure of networks and how information spreads across them. Since CSS benefits from the large availability of data from online social networks, it is attracting researchers in ever-increasing numbers as it allows for the study of mass social dynamics at an unprecedented level of resolution. Along this path, recent studies have pointed out several important results ranging from social contagion (37; 38; 39) up to information diffusion (40; 41), passing through the virality of false claims (30; 31). A wide literature branch is also devoted to understanding the spread of rumors and behaviors focusing on structural properties of social networks (30; 31; 39; 42) .

In (42) Damon Centola investigates the effects of topology on diffusion showing that network structure has a significant effect. Moreover, he observes that the behavior spread farther and faster across clustered-lattice networks than across corresponding random networks. In (39) authors use data of about 900M users on Facebook and analyze the process of contagion in the social network. They found that the probability of contagion is tightly controlled by the number of connected components in an individual's contact neighbourhood, rather than by the actual size of the neighbourhood. Thus, the chance of a user to adopt a new idea does not depend on the number of friends already holding it, but on the number of different social groups those friends hold membership in. In other words, the spread of ideas does not depend on the number of people, but on the variety of people holding them.

A key factor in identifying true contagion in social network is to distinguish between peer-to-peer influence and *homophily*: in the first case, a node influences or causes outcomes in its neighbours, while in the second one dyadic similarities between nodes create correlated outcome patterns among neighbours that merely mimic viral contagions without direct causal influence (38). In (37) authors developed an estimation framework to distinguish influence and homophily effects in dynamic networks and found that homophily explains more than 50% of the perceived behavioral contagion. These results become crucial for understanding the mechanisms behind contagions in networks and how to propagate or combat them.

Furthermore, another family of phenomena is represented by the emotional contagion, which may manifest responses that are either similar (e.g., when smiles elicit smiles) or complementary (e.g., when a fist raised in anger causes fear) (43); emotional states may be transferred to others, letting people experience the same emotions without realising it. In particular, results presented in (44) indicate that emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks. In (45) authors analyze Twitter data and measure the spatio-temporal sentiment towards a new vaccine, finding a strong correlation between sentiments expressed online and vaccination rates by region. Moreover, they show that information flows more often between users who share the same sentiment. Several studies have also tried to measure the effects of social influence online (42; 46; 47; 48). In a report on Science (49) Aral and Walker show that the propagation of behaviors of 1.3M of Facebook users is determined by the joint combinations of influence, susceptibility, and spontaneous adoption. Indeed, influential individuals are less susceptible to influence than noninfluential individuals, which suggests that influential individuals with influential friends could be crucial for the spread of information on online social networks. In a 61M Facebook users experiment during the 2010 US congressional elections (50), authors show that political mobilization messages directly influenced the voting behavior of people, as well as their political self-expression and information seeking.

Discovering and sharing information on online social networks may lead to the formation of cascades of reshares between users and thus reach a large number of individuals. A growing branch of literature focused on analyzing and characterizing such cascades, so as to discover the driving forces behind the popularity of contents. Cascades have been largely explored in different settings, such as blogging (40; 51; 52), e-mails (53; 54) and social sites e.g., Twitter (41; 55). Recent works (30; 31) also consider the anatomy and predictivity of large Facebook cascades, showing that rumor cascades run deeper in the social network than reshare cascades in general (32). Many other papers focus on the prediction of the popularity of a certain piece of content and propose rich sets of features, varying from the content, to the poster/resharer features, up to the structure and temporal properties of the cascade. In parallel, a more cautious train of thought stresses the rarity of cascades (56) and argues that their future trajectory may be unpredictable (48; 57).

Nonetheless, network structure is often not enough to understand certain dynamics. In (58) authors analyze the behavior of a large movement on Facebook. In March of 2013 3M users changed their profile picture to one of an equal sign to express their support of same-sex marriage. Authors find that, even if the number of friends played a role in the adoption of the new picture, the same did demographic and individual characteristics. Moreover, it has to be considered that someone having many friends who have changed their profile pictures is likely to be the kind of person who would join the movement, as social links are mainly created by homophily. Indeed, authors point out the importance to distinguish between influence and susceptibility in this kind of diffusion processes.

The role of social networks in information diffusion has been largely explored. In (59) researchers study 253M users on Facebook, finding that those exposed are more likely to spread information, and do so more rapidly. Furthermore, they show that it is the abundance of weak ties – i.e., people in one's *extended network* – that is responsible for the propagation of novel information.

2.2 Echo Chambers and Misinformation

The ever-increasing number of people resorting to online social networks for news and information raises important issues with respect to the creation of the so-called *echo chambers*, an enclosed system where users are exposed only to information from people having similar opinions (60). On the other hand, many argue that the formation of *filter bubbles* – where only ideologically appealing information is available – is directly related to the algorithms used to rank contents (61). Speaking of this, in (62) Facebook research scientists quantify exactly how much individuals could be exposed to ideologically diverse news and information in social media. In particular, they analyze the interaction of 10.1M users with socially shared news, finding that individual's choice about what to consume has an effect stronger than that of Facebook's News Feed algorithm in limiting the exposure to cross-cutting content.

Undoubtedly, the selective exposure to specific content facilitates the aggregation of users in echo chambers, wherein external and contradicting versions are ignored (63). Such a contest is crucial for the spread of unsubstantiated rumors; indeed, the main driver for the popularity of unverified contents becomes the *confirmation bias* – i.e., the tendency to search and interpret information in a way that it is coherent with one's beliefs or convinctions (64).

In addition, the massive diffusion of socio-technical systems and microblogging platforms has created a direct path from producers to consumers of content and changed the way in which users get informed, debate, and form their opinions. On the web the paradigm of content production and consumption is particularly disintermediated; everyone is able to produce and share contents without the mediation of an expert. This lack, especially on complex issues, might encourage speculation, rumors, and mistrust. Pages about global conspiracy, chem-trails, UFO, reptilians, or the link between vaccines and autism, proliferate on social networks, creating and promoting alternative narratives often in contrast to the mainstream one. Therefore, misinformation online is pervasive and difficult to correct.

To face the issue, several algorithmic-driven solutions have been proposed e.g., Google is developing a trustworthiness score to rank query results (65). Similarly, Facebook has proposed a community-driven approach where users can flag false contents to correct the News Feed algorithm (66). More precisely, the update to News Feed reduces the diffusion of posts reported as hoaxes and adds an annotation to posts that have received many of these types of reports to warn other users on Facebook. However, the matter is controversial, because it raises fears that the free circulation of content may be threatened and that the proposed algorithms may be inaccurate or ineffective (67).

On the other hand, the diffusion of unreliable contents might lead to confuse unverified stories with their satirical counterparts. Indeed, it has been noticed the proliferation of satirical pages producing demential imitation of conspiracy theses. In fact, there is a large variety of groups, known as *trolls*, behind the creation of Facebook pages as a caricatural version of conspiracy news. Their activities range from controversial comments and satirical posts mimicking alternative news sources, to the fabrication of purely fictitious statements, heavily unrealistic and sarcastic. Sometimes, these memes become viral and are used as evidence in online debates from political activists. Simultaneously, it has also been observed the rapid spread of blogs and pages devoted to debunk false claims, namely *debunkers*.

Such a scenario makes crucial the quantitative understanding of the social determinants related to content selection, information consumption, and beliefs formation and revision. In (68) authors study how 2.3M of Facebook users consumed different information at the edge of political discussion and news during the Italian electoral competition of 2013, showing that the social response is not affected by the topic nor by the qualitative nature of the information. Indeed, in (69) authors investigate how information related to very distinct narratives – i.e., mainstream scientific and conspiracy news – is consumed and shapes communities on Facebook, showing the emergence of polarized communities around distinct types of contents. Moreover, they find that usual consumers of conspiracy news result to be extremely focused and self-contained on their specific contents. Indeed, such a polarized structure facilitates the reinforcement and the selection of contents by confirmation bias. Based on these results, the researchers have considered to verify the effects of debunking campaigns aiming at correcting the spreading of false information on social media. In (70) they show that usual consumers of conspiracy news, when exposed to debunking news, are more prone (30%) to continue interacting with conspiracy-like information than those not exposed. In other words, trying to persuade a conspiracy user to let her beliefs fall causes exactly the opposite effect. Furthermore, by measuring the response to the injection of false information (parodistic imitations of alternative stories) they find that users prominently interacting with alternative information sources – i.e. more exposed to unsubstantiated claims – are more prone to interact with intentional and parodistic false claims (71). Indeed, homophily and polarization may be the key metrics to identify the communities of a social network where false or misleading rumors are more likely to spread (72).

Moreover, in (73) we show that the size of the echo chambers influences the size of Facebook spreading cascades. In addition, when focusing on the emotional dynamics inside and between the two echo chambers, we find that the sentiment of users on science and conspiracy pages tends to be negative, and is more and more negative when the discussion becomes longer or users activity on the social network increase (74). In particular, the discussion degenerates when the two polarized communities interact with one another.

Thus far we have focused on the behavior of the echo chambers seen from the outside, showing that contents are selected by confirmation bias and other information is ignored or rejected. Going inside the conspiracy echo chamber (75), we find that the topics belong to four main categories: *Diet*, *Environment*, *Geopolitics*, and *Health*. Moreover, we show that the more a user is active, the more he is likely to span all categories i.e., once inside a conspiracy narrative users tend to embrace the overall corpus.

Finally, we investigate the effectiveness of debunking on $54M$ users of Facebook US (36) . Our findings confirm the existence of echo chambers where users interact primarily with either conspiracy-like or scientific pages. Both groups interact similarly with the information within their echo chamber. By examining $50K$ posts we find that attempts at debunking are largely ineffective. For one, only a small fraction of usual consumers of unsubstantiated information interact with the posts. Furthermore, we show that those few are often the most committed conspiracy users and rather than internalizing debunking information, they often react to it negatively. Indeed, after interacting with debunking posts, users retain, or even increase, their engagement within the conspiracy echo chamber.

CHAPTER

THREE

EMOTIONAL DYNAMICS IN THE AGE OF MISINFORMATION

On online social media users tend to aggregate around specific contents and to form polarized group. Thus it might be supposed that users become more and more extreme in their beliefs and reinforce their initial views after discussing online. To investigate such an hypothesis, in this chapter we explore the emotional contagion of online collective debates around specific types of information – i.e., science and conspiracy news[1]. In particular, focusing on polarized users – i.e., users that are mainly active only on one type of content – we analyze their sentiment when they discuss both within their echo chamber and with the opposite faction. This work provide important insights about the relative influence of like-minded people when discussing on topics related to their preferred narrative.

[1]This work was carried out in collaboration with Petra Kralj Novak, Igor Mozetič, et al. and published on PLoS ONE on September 2015. FZ conceived, designed, and performed the experiments; analyzed the data; contributed to writing and reviewing the manuscript. See (74) for further details.

3.1 Introduction

People online get informed, discuss and shape their opinions (76; 77; 78). Indeed, microblogging platforms such as Facebook and Twitter allow for the direct and disintermediated production and consumption of contents (32; 79; 80; 81). The information heterogeneity might facilitate users selective exposure to specific content and hence the aggregation in homophilous communities (47; 72; 82; 83; 84; 85; 86; 87). In such echo-chambers users interaction with different narratives is reduced and the resulting debates are often polarized (misinformation) (68; 69; 70; 71; 73; 75).

Unfortunately, despite the enthusiastic rhetoric about *collective intelligence* (5; 88; 89), the direct and undifferentiated access to the knowledge production process is causing opposite effects – e.g., the recent case of Jade Helm 15 (90) where a simple military exercise turned out to be perceived as the beginning of the civil war in the US. Unsubstantiated rumors often jump the credulity barrier and trigger naive social responses. To an extent that, recently, the World Economic Forum labeled *massive digital misinformation* as one of the main threats to our society. Individuals may be uninformed or misinformed, and the debunking campaigns against unsubstantiated rumors do not seem to be effective (8).

Indeed, the factors behind the acceptance of a claim (whether substantiated or not) may be altered by normative social influence or by the coherence with the system of beliefs of the individual (91; 92; 93; 94; 95), making the preferential driver of contents the *confirmation bias* – i.e., the tendency to select and interpret information coherently with one's system of beliefs.

In (68; 69; 71) it has been pointed out that the more users are exposed to unsubstantiated rumors, the more they are likely to jump the credulity barrier. Recent studies (96; 97) pointed out that reading comments affects the perception of the topic and, thus, the discussion.

In this work we analyze a collection of *conspiracy* and *scientific* news sources in the Italian Facebook over a time span of four years. The main distinctive feature of the two categories of pages is the possibility to verify the reported content. Scientific news are generally fact-checked and are the results of a peer review process. Conversely, conspiracy news are generally partial information about a secret plot. We identify pages diffusing conspiracy news – i.e., pages promoting contents *neglected* by main stream media and scientific pages – aiming at diffusing scientific results. To have an exhaustive list of pages, we define the space of our investigation with the help of Facebook groups very active in debunking conspiracy stories and unverified rumors (*Protesi di Complotto*, *Che vuol dire reale*, *La menzogna diventa verità e passa alla storia*).

We target emotional dynamics inside and across content 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 annotate a substantial sample of comments, and then build a Support Vector Machine (SVM (98)) classification model. The model is then applied to associate each comment with one sentiment value: negative, neutral, or positive. The sentiment is intended to express the emotional attitude of Facebook users when posting comments.

Although other studies apply sentiment analysis to social media (99; 100; 101; 102), our work is the first linking the interplay between communities emerging around shared narratives and specifically addressing the emotional dynamics with respect to misinformation spreading. Indeed, this work provides important insights toward the understanding of the social factors behind contents consumption and the formation of polarized and homophilous clusters with a specific interest in conspiracy-like information.

We focus on the emotional behavior of about 280K Facebook Italian users and through a thorough quantitative analysis, we find that the sentiment on conspiracy pages tends to be more negative than that on science pages. In addition, by focusing on polarized users – i.e., users mainly exposed to one specific content type (science or conspiracy) – we capture an overall increase of the negativity of the sentiment. According to our results, the more active polarized users are, the more they tend to be negative, both on science and conspiracy. Furthermore, the sentiment of polarized users is negative also when they interact with one another. Also, as the number of comments increases – i.e., the discussion turns longer – the sentiment is more and more negative.

3.2 Results and Discussion

3.2.1 Sentiment Classification

Emotional attitude towards different topics can be roughly approximated by the sentiment expressed in texts. It is difficult to exactly formalize the sentiment measures since there are often disagreements between humans, and even individuals are not consistent with themselves.

In this study, as is often in the sentiment analysis literature (103), we have approximated the sentiment with an ordinal scale of three values: *negative* ($-$), *neutral* (0), and *positive* ($+$). Even with this rough approximation, and disagreements on single cases, it turns out that on a large scale, when one deals with thousands of sentiment assignments, the aggregated sentiment converges to stable values (104).

Our approach to automatic sentiment classification of texts is based on supervised machine learning. There are four steps: (i) a sample of texts is manually annotated with sentiment, (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.

We have collected over one million of Facebook comments. About 20K were randomly selected for manual annotation. We have engaged 22 native Italian speakers, active on Facebook, to manually annotate the comments by sentiment. The annotation is supported by a web-based platform Goldfinch[2] and was accomplished in two months. About 20% of the comments were intentionally duplicated, in order to measure the mutual (dis)agreement of human annotators.

[2]provided by Sowa Labs: `http://www.sowalabs.com`

There are several measures to evaluate the performance of classification models and the inter-annotator agreement. We argue that the latter provides an upper bound that the best classification model can achieve. In practice, however, different learning algorithms have various limitations, and, most importantly, only a limited amount of training data is available. In order to compare the classifier performance to the inter-annotator agreement, we have selected three measures which are applied to evaluate both, performance and agreement: $Accuracy$, $\overline{F_1}$, and $Accuracy \pm 1$. Exact definitions are in the Methods section, here we just briefly summarize them. $Accuracy$ is the fraction of correctly classified examples for all three sentiment classes – no ordering between the classes is taken into account, and all three are treated equally. F_1 is the harmonic mean of precision and recall for a selected class. $\overline{F_1}(-,+)$ is the average of F_1 for the negative and positive class only, ignoring the neutral class. It is a standard measure of performance for sentiment classification (105). The idea is that the misclassification of neutral sentiment can be ignored as it is less important then the extremes, i.e., negative or positive sentiment (however, it still affects their precision and recall). $Accuracy \pm 1$ (an abbreviation for *Accuracy within 1*) completely ignores the neutral class. It counts as errors just the negative sentiment examples predicted as positive, and vice versa. It takes into account the fact that the neutral class is between the negative and the positive, and tolerates misclassifications within neighbouring classes.

Table 1 gives the evaluation results. In the case of the inter-annotator agreement, 3,262 examples were labeled twice by two different annotators, and measures assess their agreement. In the case of a sentiment classifier evaluation, we applied 10-fold cross-validation. The results in Table 1 are the average of 10 classifiers, with 95% confidence interval. One can see that the average classifier has reached a performance close to human agreement. In terms of extreme errors, i.e., $1 - Accuracy \pm 1$ the performance of the classifier is as good as the agreement between the annotators. However, in terms of $Accuracy$ and $\overline{F_1}$, there is still some room for improvement. We speculate that the main reason for the gap is a relatively low number of annotated examples. Based on our experience in training SVM classifiers in other domains (such as stock market, elections, generic tweets, etc.), we estimate that about 50,000 to 100,000 training examples are needed to reach the level of the inter-annotator agreement.

	Annotator agreement	**Sentiment classifier**
No. of testing examples	$3,262$	$19,642$
$Accuracy(-, 0, +)$	72.0%	$64.8 \pm 1.1\%$
$\overline{F_1}(-, +)$	73.3%	$65.5 \pm 1.0\%$
$Accuracy \pm 1(-, +)$	97.2%	$97.0 \pm 0.3\%$

Table 1: Comparison of the inter-annotator agreement and classifier performance over three evaluation measures. The results for an average sentiment classifier are from 10-fold cross-validation, with 95% confidence interval.

Fig. 2 gives the distribution of sentiment values after applying the classification model to the entire set of over one million comments. We assume that the sentiment values are ordered, and that the difference from the neutral value to both extremes, negative and positive, is the same. Thus one can map the sentiment values from ordinal to a real-valued interval $[-1, +1]$. The mean sentiment over the entire set is -0.34, prevailingly negative.

3.2.2 Sentiment on Science and Conspiracy Posts

The sentiment analysis and classification task allowed us to associate each comment of our dataset to a sentiment value – i.e., -1 if *negative*, 0 if *neutral*, and 1 if *positive*. Taking all the comments of science and conspiracy posts, we can simply divide them into negative, neutral and positive (Fig. 3, *left*), and analyze their proportions. We find that 70% of comments on science pages is neutral or positive, differently from conspiracy pages (51%). Moreover, comments on science pages are twice as positive (20%) than those on conspiracy pages (10%).

To measure the effect induced on users by a post, we compute the average sentiment of all its comments. We grouped posts sentiment by

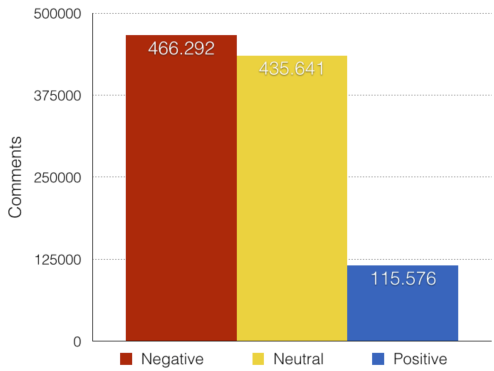


Figure 2: Sentiment distribution over the entire set of one million comments.

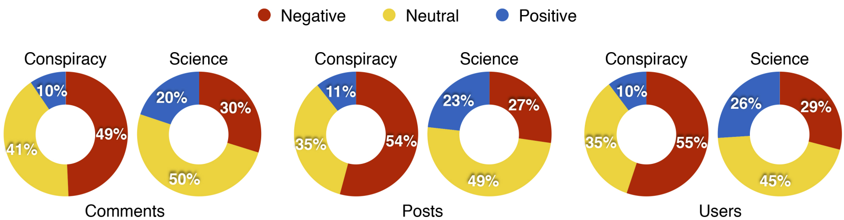


Figure 3: Proportions of negative, neutral and positive comments *(left)*, posts *(center)*, and users *(right)* both on science and conspiracy pages.

defining three thresholds in order to equally divide the space; in particular, we say a post to be *negative* if the average sentiment $\in [-1, -0.3]$, *neutral* if $\in (-0.3, 0.3)$, and *positive* if $\in [0.3, 1]$. Fig. 3 *(center)* shows the aggregated sentiment of science and conspiracy posts. Notice that the sentiment of conspiracy posts is mainly negative (54%), differently from science posts, for which the negative sentiment represents only the 27%. On the other hand, it is twice as positive for science posts (23%) than for conspiracy posts (11%).

When focusing on users, the approach is analogous. We define the sentiment of a user as the mean of the sentiment of all her comments. The mean sentiment for each user is then classified as negative, neutral, or positive by means of the same thresholds used for posts. Fig. 3 *(right)* shows the aggregated sentiment both for science and conspiracy users. We find that the sentiment of users commenting on conspiracy pages is mainly negative (55%), while the sentiment of a small fraction of users (10%) is positive. On the contrary, the sentiment of users commenting on science pages is particularly neutral (45%), and negative only for 29% of users. Almost the same percentage (26%) is represented by positive sentiment.

3.2.3 Sentiment and Virality

Now we focus on the interplay between the virality of a post and its generated sentiment. In particular we want to understand how the sentiment varies for increasing levels of comments, likes, and shares. Notice that each of these actions has a particular meaning (106; 107; 108). A *like* stands for a positive feedback to the post; a *share* expresses the will to increase the visibility of a given information; and a *comment* is the way in which online collective debates take form around the topic promoted by posts. Comments may contain negative or positive feedbacks with respect to the post. Fig. 4 shows the aggregated sentiment of a post as a function of its number of comments *(top)*, likes *(center)*, and shares *(bottom)* both for science *(left)* and conspiracy *(right)* posts. The sentiment has been regressed w.r.t. the logarithm of the number of comments (resp., likes, shares). We do not show confidence intervals, since they are defined (C.I. 95%) as $\bar{X} \pm S.E. = \bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$ and when $n \to \infty$, $S.E. = 0$. We notice that the sentiment decreases both for science and conspiracy when the number of comments of the post increases. However, we also note that it becomes more positive for science posts when the number of likes and shares increase, differently from conspiracy posts.

To assess the direct relationship between the number of comments and the negativity of the sentiment, a randomization test was performed. In particular, we took all the comments of science (resp., conspiracy) posts and randomly reassigned the original sentiments. Then, we regressed the sentiment w.r.t. the number of comments and compared the

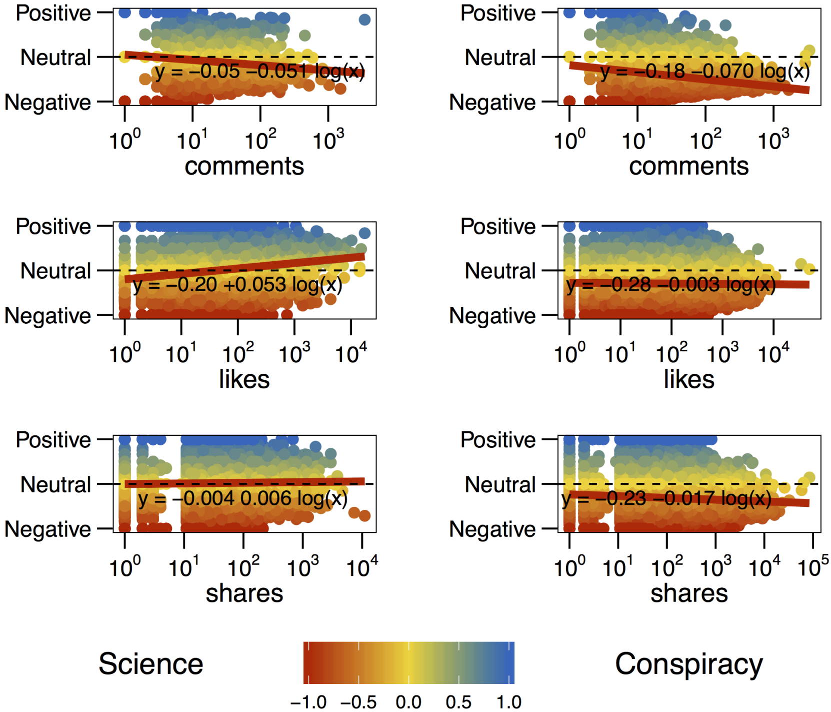


Figure 4: Sentiment and post consumption. Aggregated sentiment of posts as a function of their number of comments, likes, and shares, both for science *(left)* and conspiracy *(right)*. Negative (respectively, neutral, positive) sentiment is denoted by red (respectively, yellow, blue) color. The sentiment has been regressed w.r.t. the logarithm of the number of comments/likes/shares.

obtained slope with the one shown in Fig. 4 *(top)*. Over 10K randomized tests, the obtained slope was always greater than the original one. More precisely, while the slope for the original comments for Science is equal to -0.051 (resp., -0.070 for Conspiracy), the quantiles of the distribution of the slopes in the randomized test are: $Q_0 = -0.010$, $Q_1 = -0.002$, $Q_2 = -0.00002$, $Q_3 = 0.002$, $Q_4 = 0.010$ (resp., $Q_0 = -0.004$, $Q_1 = -0.0008$, $Q_2 = -0.000004$, $Q_3 = 0.0008$, $Q_4 = 0.005$, for Conspiracy). Therefore, given that the negative relationship between the sentiment and the length of the discussion disappears when the comment sentiments are randomized, we conclude that the length of the discussion is a relevant dimension when considering the negativity of the sentiment.

Summarizing, we found that both comments and posts, as well as users of conspiracy pages tend to be much more negative than those of science pages. Interestingly, the sentiment becomes more and more negative when the number of comments of the post increases – i.e., the discussion becomes longer – both on science and conspiracy pages. However, differently from conspiracy posts, when the number of likes and shares increases, the aggregated sentiment of science posts becomes more and more positive.

3.2.4 Sentiment and Users Activity

In this section we aim at understanding more in depth how the sentiment changes with respect to users' engagement in one of the two communities. Previous works (69; 70; 71) showed that the distribution of the users activity on the different contents is highly polarized. Therefore we now want to focus on the sentiment of polarized users. More precisely, 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 (for further details about the effect of the thresholding refer to the Methods Section).

Therefore, we take all polarized users having commented at least twice, i.e., $14,887$ out of $33,268$ users polarized on science and $67,271$ out of $135,427$ users polarized on conspiracy. Fig. 5 shows the Probability Density Function (PDF) of the mean sentiment of polarized users with at least two comments. In Table 2 we compare the mean sentiment of all users and polarized users having commented at least twice. Our results show that the overall negativity increases w.r.t. all users, such a feature is more evident on the conspiracy side.

We now want to investigate how the mean sentiment of a user changes with respect to her commenting activity – i.e., when her total number

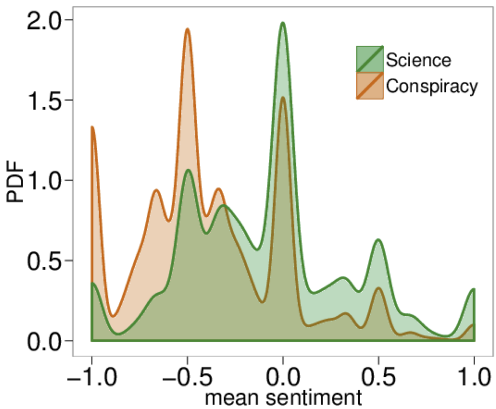


Figure 5: Sentiment and polarization. Probability Density Function (PDF) of the mean sentiment of polarized users having commented at least twice, where -1 corresponds to negative sentiment, 0 to neutral and 1 to positive.

	Science		Conspiracy	
Sentiment	**All users**	**Polarized**	**All users**	**Polarized**
Negative	29%	34%	55%	66%
Neutral	45%	46%	35%	27%
Positive	26%	20%	10%	7%

Table 2: Mean sentiment of all users and polarized users having commented at least twice.

of comments increases. In Fig. 6 we show the mean sentiment of polarized users as a function of their number of comments. The more active a polarized user is, the more she tends toward negative values both on science and conspiracy posts. The sentiment has been regressed w.r.t. the logarithm of the number of comments. Interestingly, the sentiment of science users decreases faster than that of conspiracy users. We performed a randomization test taking all comments on both categories and then randomly reassigning the original sentiments. Then, we regressed the sentiment w.r.t. the number of comments and com-

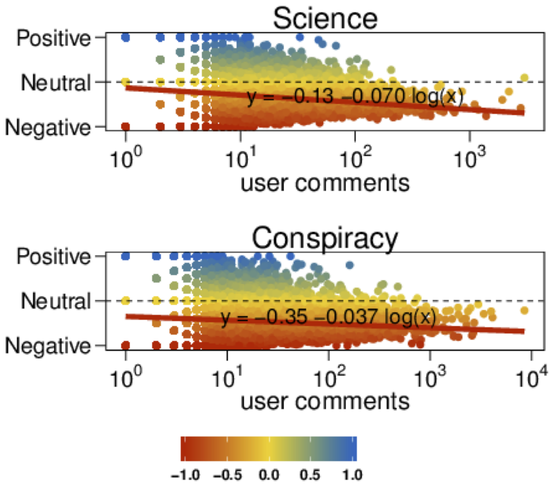


Figure 6: Sentiment and commenting activity. Average sentiment of polarized users as a function of their number of comments. Negative (respectively, neutral, positive) sentiment is denoted by red (respectively, yellow, blue) color. The sentiment has been regressed w.r.t. the logarithm of the number of comments.

pared the obtained slope with the one shown in Fig. 6. The obtained slope over 10K randomized tests was always greater than the original one. In particular, while the slope for the original comments for Science is equal to -0.070 (resp., -0.037 for Conspiracy), the quantiles of the distribution of the slopes in the randomized test are: $Q_0 = -0.006$, $Q_1 = -0.001$, $Q_2 = 0.00001$, $Q_3 = 0.001$, $Q_4 = 0.006$ (resp., $Q_0 = -0.003$, $Q_1 = -0.0005$, $Q_2 = 0.00001$, $Q_3 = 0.0005$, $Q_4 = 0.003$, for Conspiracy). Therefore users activity is a relevant dimension when considering the value of the sentiment, which is more and more negative on both categories when the users activity increases.

3.2.5 Interaction Across Communities

In this section we aim at investigating the sentiment when usual consumers of science and conspiracy news meet. To do this we pick all posts representing the arena where the debate between science and conspiracy users takes place. In particular, we select all posts commented at least once by both a user polarized on science and a user polarized on conspiracy. We find $7,751$ such posts (out of $315,567$) – reinforcing the fact that the two communities of users are strictly separated and do not often interact with one another.

In Fig. 7 we show the proportions of negative, neutral, and positive comments (*left*) and posts (*right*). The aggregated sentiment of such posts is slightly more negative (60%) than for general posts (54% for conspiracy, 27% for science, see Fig. 3). When focusing on comments, we have similar percentages of neutral (42%) and negative (48%) comments, while a small part (10%) is represented by positive comments. We want to understand if the sentiment correlates with the length of the discussion. Hence, we analyze how the sentiment changes when the number of comments of the post increases, as we previously did for *general* posts (Fig. 4). Fig. 8 shows the aggregated sentiment of such posts as a function of their number of comments. Clearly, as the number of comments increases – i.e., the discussion becomes longer – the sentiment is more and more negative. Moreover, comparing with Fig. 4, when communities interact with one another, posts show a higher concentration of negative sentiment. Also in this case we performed a randomization test taking all the comments and randomly reassigning the original sentiments. Then, we regressed the sentiment w.r.t. the number of comments and compared the obtained slope with the one shown in Fig. 7. Over 10K randomized tests, the obtained slope was always greater than the original one. In particular, while the slope for the original comments is equal to -0.048, the quantiles of the distribution of the slopes in the randomized test are: $Q_0 = -0.009$, $Q_1 = -0.002$, $Q_2 = 0.00004$, $Q_3 = 0.002$, $Q_4 = 0.009$. Therefore, we conclude that the length of the discussion does affect the negativity of the sentiment.

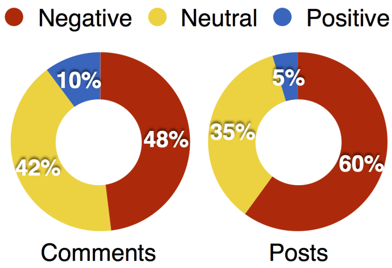


Figure 7: Sentiment between communities. Proportions of negative, neutral, and positive comments *(left)* and posts *(right)* of all the posts commented at least once by both a user polarized on science and a user polarized on conspiracy.

3.3 Methods

3.3.1 Ethics Statement and Data Collection

The entire data collection process has been carried out exclusively through the Facebook Graph API (109), which is publicly available, and we used only public available data – users with privacy restrictions are not included in the dataset – for the analysis (according to the specification settings of the API). The pages from which we downloaded data are public Facebook entities (can be accessed by anyone). Users' content contributing to such pages is also public unless the user's privacy settings specify otherwise and in that case it is not available to us.

We identified two main categories of pages: conspiracy news – i.e. pages promoting contents neglected by main stream media – and science news. The first category includes all pages diffusing conspiracy information – pages which disseminate controversial information, most often lacking supporting evidence and sometimes contradictory to the official news (i.e., conspiracy theories). The second category is that of scientific dissemination, including scientific institutions and scientific press having the main mission to diffuse scientific knowledge. Note that we do not focus on the truth value of information but rather on the possibil-

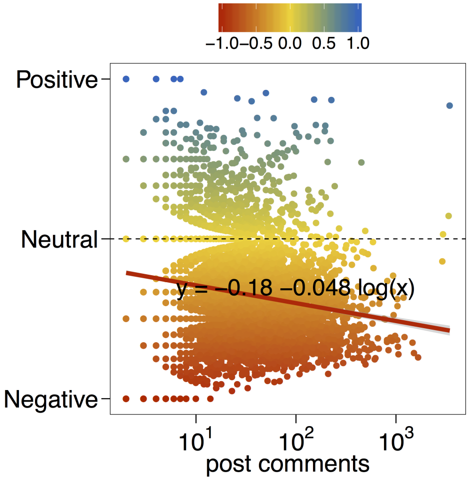


Figure 8: Sentiment and discussion. Aggregated sentiment of posts as a function of their number of comments. Negative (respectively, neutral, positive) sentiment is denoted by red (respectively, yellow, blue) color.

ity of verifying the content of the page. While the latter is an easy task for scientific news – e.g., by identifying the authors of the study or if the paper passed a peer review process – it usually becomes more difficult for conspiracy-like information, if not unfeasible. We defined the space of our investigation with the help of Facebook groups very active in debunking conspiracy theses (*Protesi di Complotto*, *Che vuol dire reale*, *La menzogna diventa verità e passa alla storia*). We categorized pages according to their contents and their self description. The resulting dataset – downloaded over a timespan of four years (2010 to 2014) – is composed of 73 public Italian Facebook pages and it is the same used in (69) and (70). To the best of our knowledge, the final dataset is the complete set of all scientific and conspiracy information sources active in the Italian Facebook scenario. Table 3 summarizes the details of our data collection (see Appendix B.1 for the complete list of pages).

	Science	Conspiracy	Total
Pages	34	39	73
Posts	62,075	208,591	270,666
Likes	2,505,399	6,659,382	9,164,781
Comments	180,918	836,591	1,017,509
Shares	1,471,088	16,326,731	17,797,819
Likers	332,357	864,047	1,196,404
Commenters	53,438	226,534	279,972

Table 3: Breakdown of the Facebook dataset.

3.3.2 Classification and Annotator Agreement Measures

Our approach to sentiment classification of texts is based on supervised machine learning, where a sample of texts is first manually annotated with sentiment and then used to train and evaluate a classifier. The classifier is then applied to the whole corpus. The measures to assess the agreement between annotators and the quality of the classifier are based on coincidence and confusion matrices, respectively.

Annotators were asked to label each text with *negative* $\prec$ *neutral* $\prec$ *positive* sentiment. When two annotators are given the same text, they can either agree (both give the same label) or disagree (they give different labels). The annotators can disagree in two ways: one label is *neutral* while the other is extreme (*negative* or *positive*), or both are extreme: one *negative* and one *positive* – we call this severe disagreement. A convenient way to represent the overall (dis)agreement between the annotators is a coincidence matrix, where each text that is annotated twice appears in the table twice. Table 4 gives a generic 3×3 annotator agreement table, while the actual data are in Tables 5 and 6. All agreements are on the diagonal of the table. As the labels are ordered (*negative* $\prec$ *neutral* $\prec$ *positive*), the further the cell from the diagonal, the more severe is the error. From such a table one can calculate the annotator agreement (the sum of the main diagonal divided by the number of all the elements in the table) and the severe disagreement: the sum of top right and bottom left corners divided by the number of all the elements in the table.

To compare the predictions of a classifier to a golden standard (manually annotated data, in our case), a confusion matrix is used. Table 4 also represents a generic 3×3 confusion matrix for the (ordered) sentiment classification case. Each element $\langle x, y \rangle$ represents the number of examples from the actual class x, predicted as class y. All agreements/correct predictions are in the diagonal of the table. In the ordinal classification case, the further the cell from the diagonal, the more severe is the error.

Actual/Predicted	*Negative*	*Neutral*	*Positive*	**Total**
Negative	$\langle -, - \rangle$	$\langle -, 0 \rangle$	$\langle -, + \rangle$	$\langle -, * \rangle$
Neutral	$\langle 0, - \rangle$	$\langle 0, 0 \rangle$	$\langle 0, + \rangle$	$\langle 0, * \rangle$
Positive	$\langle +, - \rangle$	$\langle +, 0 \rangle$	$\langle +, + \rangle$	$\langle +, * \rangle$
Total	$\langle *, - \rangle$	$\langle *, 0 \rangle$	$\langle *, + \rangle$	N

Table 4: A generic 3×3 coincidence matrix/confusion matrix. An element $\langle x, y \rangle$ denotes the number of examples from the actual class x, predicted as class y.

$Accuracy$ is the fraction of correctly classified examples:

$$Accuracy = \frac{\langle -, - \rangle + \langle 0, 0 \rangle + \langle +, + \rangle}{N}$$

Accuracy within n (110) allows for a wider range of predictions to be considered correct. We use *Accuracy within 1* ($Accuracy \pm 1$) where only misclassifications from *negative* to *positive* and vice-versa are considered incorrect:

$$Accuracy \pm 1(-, +) = 1 - \frac{\langle +, - \rangle + \langle -, + \rangle}{N}$$

$\overline{F_1}(+, -)$ is the macro-averaged F-score of the positive and negative classes, a standard evaluation measure (105) used also in the SemEval competition[3] for sentiment classification tasks:

$$\overline{F_1}(+, -) = \frac{F_{1+} + F_{1-}}{2}$$

[3] `http://alt.qcri.org/semeval2015/`

F_1 is the harmonic mean of *Precision* and *Recall* for each class (111):

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Precision for class x is the fraction of correctly predicted examples out of all the predictions with class x:

$$Precision_x = \frac{\langle x, x \rangle}{\langle *, x \rangle}$$

Recall for class x is the fraction of correctly predicted examples out of all the examples with actual class x:

$$Recall_x = \frac{\langle x, x \rangle}{\langle x, * \rangle}$$

From the above tables and definitions, one can see that the annotator agreement is equivalent to $Accuracy$ and that severe disagreement is equivalent to $1 - Accuracy_{\pm 1}$. $\overline{F_1}$ has no counterpart between the annotator agreement measures, but is a standard measure in evaluation of sentiment classifiers. On the other hand, Cohen's kappa (112) is a standard measure of inter-rater agreement, but rarely used to evaluate classification models. The original Cohen's kappa is applicable to categorical (unordered) classes, and weighted kappa was devised for ordered classes. We use *Cohen's weighted kappa* (113) to compare the inter-annotator agreement and self-agreement.

3.3.3 Data Annotation

Data annotation is a process in which some predefined labels are assigned to each data point. A subset of 19,642 comments from the Facebook dataset of one million (Table 3) was selected for manual sentiment annotation and later used to train a sentiment classifier. A user-friendly web and mobile devices annotation platform Goldfinch[4] was used.

Trustworthy Italian native speakers, active on Facebook, were engaged for the annotations. The annotation task was to label each Facebook comment – isolated from its context – as *negative*, *neutral*, or *positive*.

[4] provided by Sowa Labs: `http://www.sowalabs.com/`

The guideline given to the annotators was to estimate the emotional attitude of the user when posting a comment to Facebook. The exact question an annotator should answer was: "Is the user happy (pleased, satisfied), or unhappy (angry, sad, frustrated), or neutral?" A dedicated Facebook group was formed to dispatch detailed annotation instructions, to provide a forum for discussion, and to post ongoing annotation results which stimulated the annotators to contribute. During the annotation process, which lasted for about two months, the annotator performance was monitored in terms of the inter-annotator agreement and self-agreement, based on 20% of the comments which were intentionally duplicated. No compensation, other then gratitude and personal satisfaction for contributing to interesting scientific research, was awarded.

	Negative	*Neutral*	*Positive*	**Total**
Negative	2,482	545	90	3,117
Neutral	545	1,474	277	2,296
Positive	90	277	744	1,111
Total	3,117	2,296	1,111	6,524

Table 5: A coincidence matrix for the inter-annotator agreement, excluding self-agreement.

The annotation process resulted in 19,642 sentiment labeled comments, 3,902 of them annotated twice. Out of 3,902 duplicates, 3,262 were polled twice to two different annotators and are used to assess the inter-annotator agreement, and 640 were polled twice to the same annotator and are used to asses the annotators' self-agreement. The coincidence matrices with the inter-annotator agreement and self-agreement are in Tables 5 and 6, respectively.

Note that, in a coincidence matrix, each annotated example appears twice (once for each of the two annotators), thus the matrix is symmetric. This is in contrast to a confusion matrix where one knows the ground truth, and the matrix values are the number of examples in the actual and predicted classes.

The four evaluation measures, defined above, were used to quantify

	Negative	*Neutral*	*Positive*	**Total**
Negative	486	57	6	549
Neutral	57	434	19	510
Positive	6	19	196	221
Total	549	510	221	1280

Table 6: A coincidence matrix for the annotators' self-agreement.

the inter-annotator and the annotators' self-agreement. The results are in Table 7.

	Inter-annotator agreement	**Annotators' self-agreement**
No. of overlapping examples	$3,262$	640
$Accuracy(-,0,+)$	72.0%	87.2%
$\overline{F_1}(-,+)$	73.3%	88.7%
$Accuracy{\pm}1(-,+)$	97.2%	99.1%
Cohen's weighted kappa	0.61	0.82

Table 7: Comparison of the inter-annotator and self-agreement over four evaluation measures.

3.3.4 Classification

Ordinal classification, also known as ordinal regression, is a form of multi-class classification where there is a natural ordering between the classes, but no meaningful numeric difference between them (110). We treat sentiment classification as an ordinal regression task with three ordered classes. We apply the wrapper approach, described in (114), with two linear-kernel Support Vector Machine (SVM) (98) classifiers. SVM is a state-of-the-art supervised learning algorithm, well suited for large scale text categorization tasks, and robust on large feature spaces. The two SVM classifiers were trained to distinguish the extreme classes (*negative* and *positive*) from the rest (*neutral* plus *positive*, and *neutral* plus *negative*, respectively). During prediction, if both classifiers agree, they yield the common class, otherwise, if they disagree, the assigned class is *neutral*.

The sentiment classifier was trained and tuned on the training set of 15,714 annotated comments. The comments were processed into the standard Bag-of-Words (BoW) representation, with the following settings: lemmatized BoW include unigrams and bigrams, minimum n-gram frequency is five, TF-IDF weighting, no stop-word removal, and normalized vectors. Additional features and settings were chosen, based on the results of 10-fold stratified cross-validation on the training set: normalization of diacritical characters, url replacement, length of text, presence of upper cased words, negation (language specific), swearing (language specific), positive words from a predefined dictionary (language specific), unusual punctuation (several exclamation or question marks, ...), unusually repeated characters, happy or sad emoticons in the text, and their presence at the end of the sentence.

The trained sentiment classifier was then evaluated on a disjoint test set of the remaining 3,928 comments. The confusion matrix between the annotators (actual classes) and the classifier (predicted classes) is in Table 8. The sentiment class distribution, after applying the classifier to the whole set of one million Facebook comments, is in Fig. 2.

Actual/Predicted	*Negative*	*Neutral*	*Positive*	**Total**
Negative	1,208	501	32	1,741
Neutral	509	987	103	1,599
Positive	86	183	319	588
Total	1,803	1,671	454	3,928

Table 8: A confusion matrix of the sentiment classifier on the test set.

Another evaluation was performed by a 10-fold cross-validation on the complete set of 19,642 training examples. The confusion matrix between the annotators and the 10 classifiers is in Table 9. The averaged evaluation measures over 10 classifiers, with 95% confidence interval are in Table 1.

Actual/Predicted	*Negative*	*Neutral*	*Positive*	**Total**
Negative	5,779	2,669	302	8,750
Neutral	1,969	5,090	839	7,898
Positive	293	834	1,867	2,994
Total	8,041	8,593	3,008	19,642

Table 9: A confusion matrix of sentiment classifiers on the 10-fold cross-validated complete training set.

3.3.5 Labelling Algorithm

The labelling algorithm may be described as a thresholding strategy on the total number of users likes. Considering the total number of likes of a user L_u on both posts P in categories S and C. Let l_s and l_c define the number of likes of a user u on P_s or P_c, respectively denoting posts from scientific or conspiracy pages. Then, the total like activity of a user on one category is given by $\frac{l_s}{L_u}$. Fixing a threshold θ we can discriminate users with enough activity on one category. More precisely, the condition for a user to be labeled as a polarized user in one category can be described as $\frac{l_s}{L_u} \vee \frac{l_c}{L_u} > \theta$. In Fig. 9 we show the number of polarized users as a function of θ. Both curves decrease with a comparable rate. Fig. 10 shows the Probability Density Function (PDF) of the mean sentiment of all polarized users *(top)* and polarized users with at least five likes *(bottom)*. Note that both densities are qualitatively similar. In Fig. 11 we show the mean sentiment of polarized users as a function of the threshold θ.

3.4 Conclusions

In this work we analyzed the emotional dynamics on pages of opposite worldviews, science and conspiracy. Previous works (69; 70; 71) showed that users are strongly polarized towards the two narratives. Moreover, we found that users of both categories seem to not distinguish between verified contents and unintentional false claims. In this manuscript we focused on the emotional behavior of the same users on Facebook. In

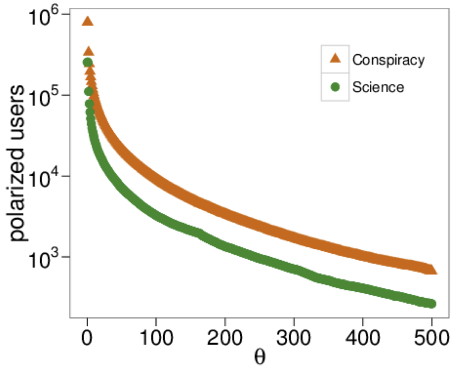


Figure 9: Polarized users and activity. The number of polarized users as a function of the thresholding value θ on the two categories.

general, we noticed that the sentiment on conspiracy pages tends to be more negative than that on science pages. In addition, by focusing on polarized users, we identified an overall increase of the negativity of the sentiment. In particular, the more active polarized users, the more they tend to be negative, both on science and conspiracy. Furthermore, the sentiment of polarized users is negative also when they interact with one another. Also in this case, as the number of comments increases – i.e., the discussion becomes longer – the sentiment of the post is more and more negative. This work provides important insights about the emotional dynamics in a disintermediated environment. Indeed, recent studies (96; 97) pointed out that reading comments of other user may affect the discussion. Our findings confirm such a phenomenon and make explicit that the longer the discussion the more negative the sentiment. In particular, discussions around conspiracy news degenerate faster than the scientific one. This latter point opens to interesting question about the quasi-religious mentality of conspiracists (115) and the way in which such an echo-chamber digests and debate news and events.

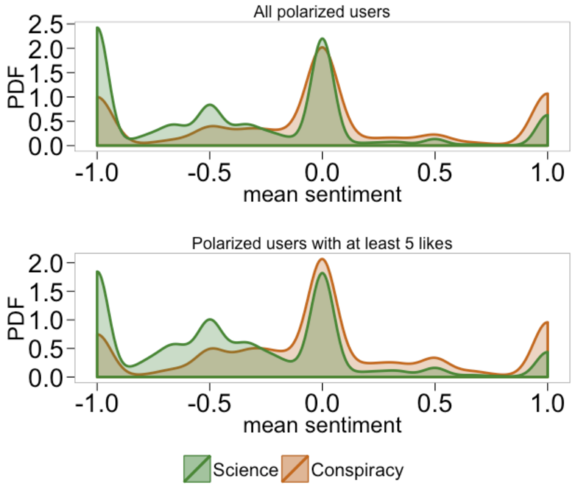


Figure 10: Sentiment of Polarized Users. Probability Density Function (PDF) of the mean sentiment of all polarized users (top) and polarized users with at least five likes, where -1 corresponds to negative sentiment, 0 to neutral and 1 to positive.

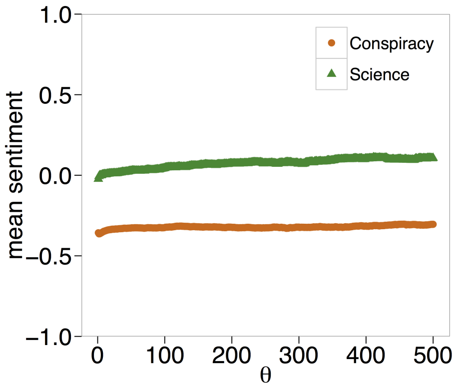


Figure 11: Sentiment and Engagement. Average sentiment of polarized users as a function of the threshold θ, i.e., the engagement degree, intended as the number of likes a polarized user put in her own category.

CHAPTER
FOUR

DEBUNKING IN A WORLD OF TRIBES

Users tend to focus on specific narratives and to join polarized groups where debating influences negatively their emotions. As a further step towards the understanding of polarization dynamics, in this chapter we want to test the response of polarized users – users interacting mainly with just one type of content – to dissenting information[1].

Focusing on users in the US conspiracy echo chamber we want to characterize their interaction with information aimed at debunking their beliefs i.e., attempts to correct unverified rumors[2].

4.1 Introduction

Socio-technical systems and microblogging platforms such as Facebook and Twitter have created a direct path from producers to consumers of

[1]This work was carried out in collaboration with Prof. Shlomo Havlin et al. and is currently under review on PLoS ONE. FZ conceived and designed the experiments; performed the analysis and interpreted the results; contributed to writing and reviewing the manuscript. See (36) for further details.

[2]The research work presented in this chapter had a profound impact on the public opinion and on the media. Based on such results, Caitlin Dewey decided to close her weekly column on the Washington Post, *What was Fake*, launched on May 2014 (116). Prof. Cass R. Sunstein, the former administrator of the White House Office of Information and Regulatory Affairs, expressed his interest in the same research on Bloomberg View, and argued that the best solution is to promote a culture of humility and openness (117).

content, changing the way users get informed, debate ideas, and shape their worldviews (47; 85; 118; 119; 120). Misinformation on online social media is pervasive and represents one of the main threats to our society according to the World Economic Forum (9; 121). The diffusion of false rumors affects public perception of reality as well as the political debate (8). Indeed, links between vaccines and autism, the belief that 9/11 was an inside job, or the more recent case of Jade Helm 15 – a simple military exercise that was perceived as the imminent threat of the civil war in the US – are just few examples of the consistent body of the collective narratives grounded on unsubstantiated information.

Confirmation bias plays a pivotal role in cascades dynamics and facilitates the emergence of echo chambers (73). Indeed, users online show the tendency a) to select information that adheres to their system of beliefs even when containing parodistic jokes; and b) to join polarized groups (122). Recently, researches have shown (68; 69; 71; 74; 75; 123) that continued exposure to unsubstantiated rumors may be a good proxy to detect gullibility – i.e., jumping the credulity barrier by accepting highly implausible theories – on online social media. Narratives, especially those grounded on conspiracy theories, play an important cognitive and social function in simplifying causation. They are formulated in a way that is able to reduce the complexity of reality and to tolerate a certain level of uncertainty (91; 92; 94). However, conspiracy thinking creates or reflects a climate of disengagement from mainstream society and recommended practices (124).

Several efforts are striving to contrast misinformation spreading from algorithmic-based solutions to tailored communication strategies (125; 126; 127; 128; 129; 130) but not much is known about their efficacy. In this work we characterize the consumption of debunking posts on Facebook and, more generally, the reaction of users to dissenting information.

We perform a thorough quantitative analysis of 54 million US Facebook users and study how they consume scientific and conspiracy-like contents. We identify two main categories of pages: conspiracy news – i.e. pages promoting contents *neglected* by main stream media – and science news. Using an approach based on (68; 69; 71), we further explore Facebook pages that are active in debunking conspiracy theses (see Section 4.3 for further details about data collection).

Notice that we do not focus on the quality of the information but rather on the possibility for verification. Indeed, it is easy for scientific news to identify the authors of the study, the university under which the study took place and if the paper underwent a peer review process. On the other hand, conspiracy-like content is difficult to verify because it is inherently based upon suspect information and is derived allegations and a belief in secrets from the public. The self-description of many conspiracy pages on Facebook, indeed, claims that they inform people about topics neglected by mainstream media and science. Pages like *I don't trust the government*, *Awakening America*, or *Awakened Citizen*, promote wide-ranging content from aliens, chem-trails, to the causal relation between vaccinations and autism or homosexuality. Conversely, science news pages – e.g., *Science*, *Science Daily*, *Nature* – are active in diffusing posts about the most recent scientific advances.

The list of pages has been built by censing all pages with the support of very active debunking groups (see Section 4.3 for more details). The final dataset contains pages reporting on scientific and conspiracy-like news. On a time span of five years (Jan 2010, Dec 2014) we downloaded all public posts (with the related lists of likes and comments) of 83 scientific and 330 conspiracy pages. In addition, we identified 66 Facebook pages aiming at debunking conspiracy theories.

Our analysis shows that two well-formed and highly segregated communities exist around conspiracy and scientific topics – i.e., users are mainly active in only one category. Focusing on users interactions with respect to their preferred content, we find similarities in the consumption of posts. Different kinds of content aggregate polarized groups of users (echo chambers). At this stage we want to test the role of confirmation bias with respect to dissenting (resp., confirmatory) information from the conspiracy (resp., science) echo chamber. Focusing on a set of $50,220$ debunking posts we measure the interaction of users from both conspiracy and science echo chambers. We find that such posts remain confined to the scientific echo chamber mainly. Indeed, the majority of likes on debunking posts is left by users polarized towards science ($\sim 67\%$), while only a small minority ($\sim 7\%$) by users polarized towards conspiracy. However, independently of the echo chamber, the sentiment expressed by users when commenting on debunking posts is mainly negative.

4.2 Results and Discussion

The aim of this work is to test the effectiveness of debunking campaigns on online social media. As a more general aim we want to characterize and compare users attention with respect to a) their preferred narrative and b) information dissenting from such a narrative. Specifically we want to understand how users usually exposed to unverified information such as conspiracy theories respond to debunking attempts.

Echo chambers

As a first step we characterize how distinct types of information – belonging to the two different narratives – are consumed on Facebook. In particular we focus on users' actions allowed by Facebook's interaction paradigm – i.e., likes, shares, and comments. Each action has a particular meaning (106). A *like* represents a positive feedback to a post; a *share* expresses a desire to increase the visibility of a given information; and a *comment* is the way in which online collective debates take form around the topic of the post. Therefore, comments may contain negative or positive feedbacks with respect to the post.

Assuming that a user u has performed x and y likes on scientific and conspiracy-like posts, respectively, we let $\rho(u) = (y - x)/(y + x)$. Thus, a user u for whom $\rho(u) = -1$ is polarized towards science, whereas a user whose $\rho(u) = 1$ is polarized towards conspiracy. We define the user polarization $\rho_{likes} \in [-1, 1]$ (resp., $\rho_{comments}$) as the ratio of difference in likes (resp., comments) on conspiracy and science posts. In Fig 12 we show that the probability density function (PDF) for the polarization of all users is sharply bimodal with most having ($\rho(u) \sim -1$) or ($\rho(u) \sim 1$). Thus, most users may be divided into two groups, those *polarized towards* *science* and those *polarized towards conspiracy*. The same pattern holds if we look at polarization based on comments rather than on likes.

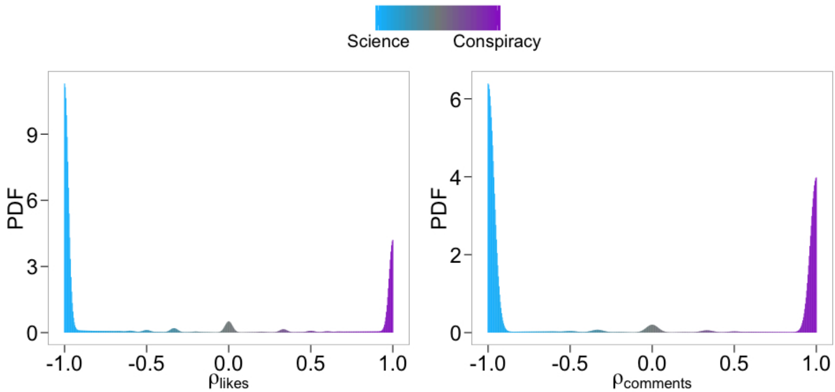


Figure 12: Users polarization. Probability density functions (PDFs) of the polarization of all users computed both on likes *(left)* and comments *(right)*.

To further understand how these two segregated communities behave, we explore how they interact with their preferred type of information. In the left panel of Fig 13 we show the distributions of the number of likes, comments, and shares on posts belonging to both scientific and conspiracy news. As seen from the plots, all the distributions are heavy-tailed – i.e, all the distributions are best fitted by power laws and all possess similar scaling parameters (see Section 4.3 for further details).

We define the persistence of a post (resp., user) as the Kaplan-Meier estimates of survival functions by accounting for the first and last comment to the post (resp., of the user). In the right panel of Fig 13 we plot the Kaplan-Meier estimates of survival functions of posts grouped by category. To further characterize differences between the survival functions, we perform the Peto & Peto (131) test to detect whether there is a statistically significant difference between the two survival functions. Since we obtain a p-value of 0.944, we can state that there are not significant statistical differences between the posts' survival functions on both science and conspiracy news. Thus, the posts' persistence is similar in

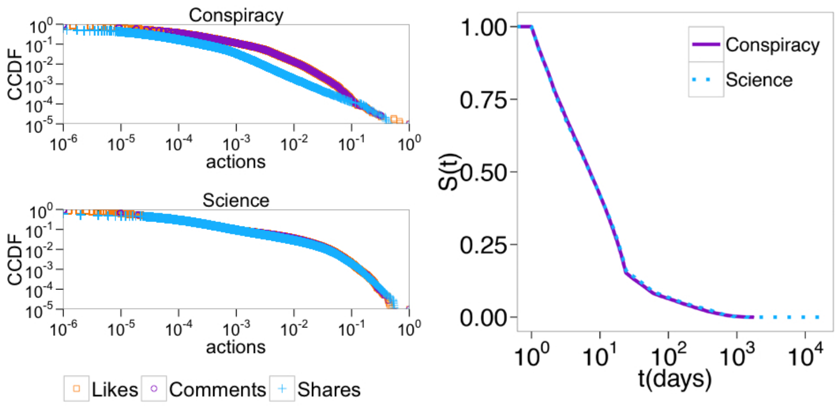


Figure 13: Posts' attention patterns and persistence. *Left panel:* Complementary cumulative distribution functions (CCDFs) of the number of likes, comments, and shares received by posts belonging to conspiracy *(top)* and scientific *(bottom)* news. *Right panel:* Kaplan-Meier estimates of survival functions of posts belonging to conspiracy and scientific news. Error bars are on the order of the size of the symbols.

the two echo chambers.

We continue our analysis by examining users interaction with different kinds of posts on Facebook. In the left panel of Fig 14 we plot the CCDFs of the number of likes and comments of users on science or conspiracy news. These results show that users consume information in a comparable way – i.e, all distributions are heavy tailed (for scaling parameters and other details refer to Section 4.3). The right panel of Fig 14 shows that the persistence of users – i.e., the Kaplan-Meier estimates of survival functions – on both types of content is nearly identical. Attention patterns of users in the conspiracy and science echo chambers reveal that both behave in a very similar manner.

In summary, contents related to distinct narratives aggregate users into different communities and consumption patterns are similar in both communities.

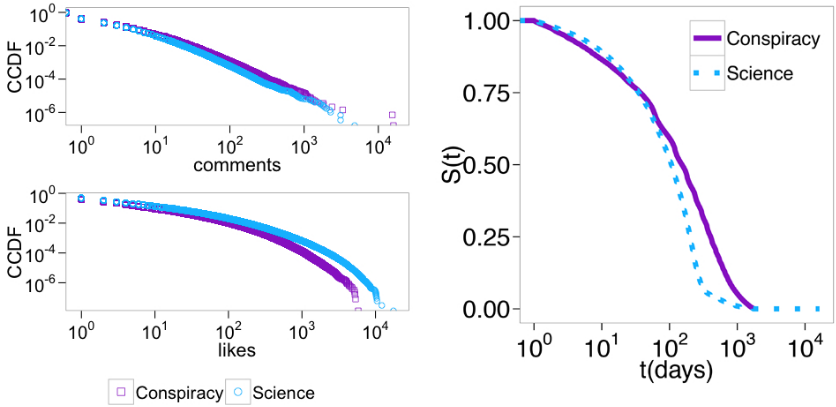


Figure 14: Users' attention patterns and persistence. *Left panel:* Complementary cumulative distribution functions (CCDFs) of the number of comments *(top)*, and likes *(bottom)*, per each user on the two categories. *Right panel:* Kaplan-Meier estimates of survival functions for users on conspiracy and scientific news. Error bars are on the order of the size of the symbols.

Response to debunking posts

Debunking posts on Facebook strive to contrast misinformation spreading by providing fact-checked information to specific topics. However, not much is known about the effectiveness of debunking to contrast misinformation spreading. In fact, if confirmation bias plays a pivotal role in selection criteria, then debunking might sound to users usually exposed to unsubstantiated rumors like something dissenting from their narrative. Here, we focus on the scientific and conspiracy echo chambers and analyze consumption of debunking posts. As a preliminary step we show how debunking posts get liked and commented according to users polarization. Notice that we consider a user to be polarized if at least the 95% of his liking activity concentrates just on one specific narrative. Fig 15 shows how users' activity is distributed on debunking posts: Left (resp., right) panel shows the proportions of likes (resp., comments) left by users polarized towards science, users polarized towards conspiracy, and not polarized users. We notice that the majority of both likes and comments is left by users polarized towards science (resp., $66,95\%$ and $52,12\%$), while only a small minority is made by users polarized towards conspiracy (resp., $6,54\%$ and $3,88\%$). Indeed, the scientific echo chamber is the biggest consumer of debunking posts and only few users usually active in the conspiracy echo chamber interact with debunking information. Out of $9,790,906$ polarized conspiracy users, just $117,736$ interacted with debunking posts – i.e., commented a debunking post at least once.

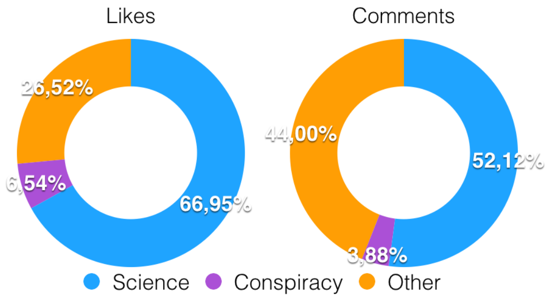


Figure 15: Users' activity on debunking posts. Proportions of likes *(left)* and comments *(right)* left by users polarized towards science, users polarized towards conspiracy, and not polarized users.

To better characterize users' response to debunking attempts, we apply sentiment analysis techniques to the comments of the Facebook posts (see Section 4.3 for further details). We use a supervised machine learning approach: first, we annotate a sample of comments and, then, we build a Support Vector Machine (SVM) (98) classification model. Finally, we apply the model to associate each comment with a sentiment value: *negative*, *neutral*, or *positive*. The sentiment denotes the emotional attitude of Facebook users when commenting. In Fig 16 we show the fraction of negative, positive, and neutral comments for all users and for the polarized ones. Notice that we consider only posts having at least a like, a comment, and a share. Comments tend to be mainly negative and such a negativity is dominant regardless users polarization.

Our findings show that debunking posts remain mainly confined within the scientific echo chamber and only few users usually exposed to un-

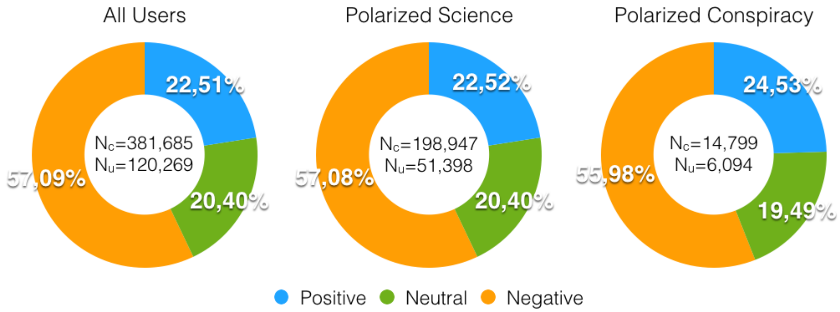


Figure 16: Users' sentiment on debunking posts. Sentiment of comments made by all users *(left)*, users polarized towards science *(center)*, and users polarized towards conspiracy *(right)* on debunking posts having at least a like, a comment, and a share.

substantiated claims actively interact with the corrections. Dissenting information is mainly ignored. Furthermore, if we look at the sentiment expressed by users in their comments, we find a rather negative environment.

Interaction with dissenting information

Users tend to focus on a specific narrative and select information adhering to their system of beliefs while they ignore dissenting information. However, in our scenario few users belonging to the conspiracy echo chamber interact with debunking information. What about such users? And further, what about the effect of their interaction with dissenting information? In this section we aim at better characterizing the consumption patterns of the few users that tend to interact with dissenting information. Focusing on the conspiracy echo chamber, in the top panel of Fig 17 we show the distinct survival functions – i.e. the probability of continuing in liking and commenting along time on conspiracy posts – of users who commented or not on debunking posts. Users interacting with debunking posts are generally more likely to survive – to pursue their interaction with conspiracy posts.

The bottom panel of Fig 17 shows the CCDFs of the number of likes

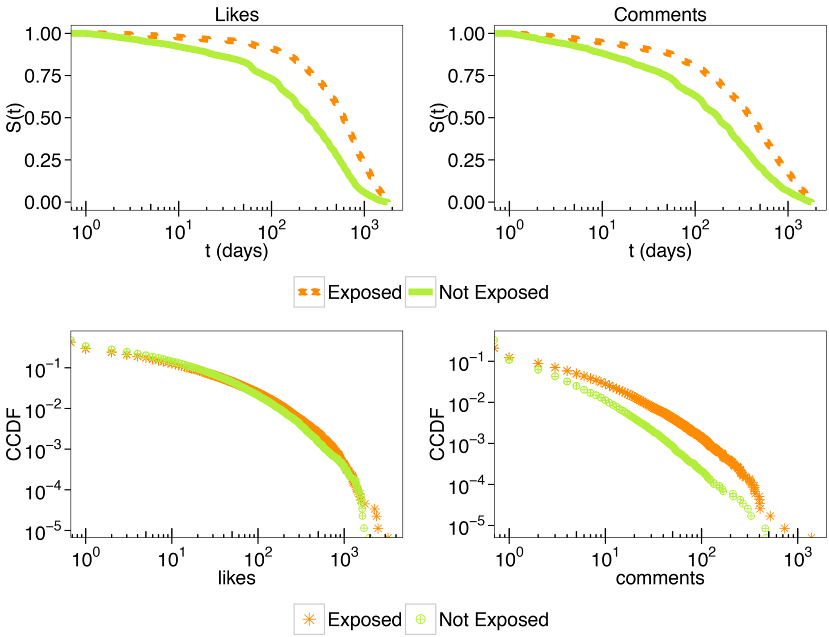


Figure 17: Interaction with debunking: survival functions and attention patterns. *Top panel:* Kaplan-Meier estimates of survival functions of users who interacted (exposed) and did not (not exposed) with debunking. Users persistence is computed both on their likes *(left)* and comments *(right)*. *Bottom panel:* Complementary cumulative distribution functions (CCDFs) of the number of likes *(left)* and comments *(right)*, per each user exposed and not exposed to debunking.

and comments for both type of users. The Spearman's rank correlations coefficient between the number of likes and comments for both type of users are very similar: $\rho_{exp} = 0.53$ (95% c.i. [0.529, 0.537]); $\rho_{not\_exp} = 0.57$ (95% c.i. [0.566, 0.573]). However, we may observe that users who commented to debunking posts are slightly *more* prone to comment in general. Thus, users engaging debates with debunking posts seems to be those few who show a higher commenting activity overall.

To further characterize the effect of the interaction with debunking posts, as a secondary step, we perform a comparative analysis between the users behavior before and after they comment on debunking posts. Fig 18 shows the liking and commenting rate – i.e, the average number of likes (or comments) on conspiracy posts per day – before and after the first interaction with debunking. The plot shows that users' liking and commenting rates increase after commenting. To further analyze the effects of interaction with the debunking posts we use the Cox Proportional Hazard model (132) to estimate the hazard of conspiracy users exposed to – i.e., who interacted with – debunking compared to those not exposed and we find that users not exposed to debunking are 1.76 times more likely to stop interacting with conspiracy news (see Section 4.3 for further details).

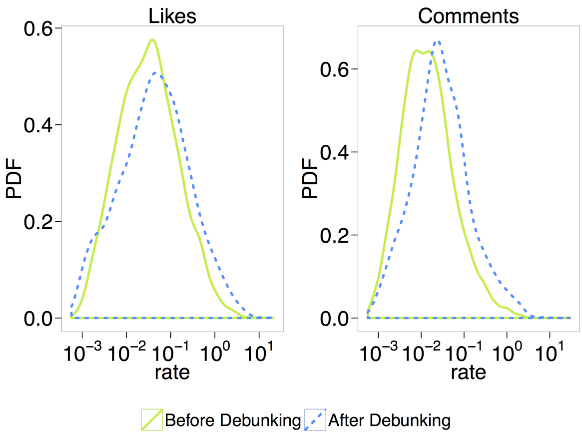


Figure 18: Interaction with debunking: comments and likes rate. Rate – i.e., average number, over time, of likes *(left)* (resp., comments *(right)*) on conspiracy posts of users who interacted with debunking posts.

Conclusions

Users online tend to focus on specific narratives and select information adhering to their system of beliefs. Such a polarized environment might foster the proliferation of false claims. Indeed, misinformation is pervasive and really difficult to correct. To smooth the proliferation of unsubstantiated rumors major corporations such as Facebook and Google are studying specific solutions. Examining the effectiveness of online debunking campaigns is crucial for understanding the processes and mechanisms behind misinformation spreading. In this work we show the existence of social echo chambers around different narratives on Facebook in the US. Two well-formed and highly segregated communities exist around conspiracy and scientific topics – i.e., users are mainly active in only one category. Furthermore, by focusing on users interactions with respect to their preferred content, we find similarities in the way in which both forms of content are consumed.

Our findings show that debunking posts remain confined within the scientific echo chamber mainly and only few users usually exposed to unsubstantiated claims actively interact with the corrections. Dissenting information is mainly ignored and, if we look at the sentiment expressed by users in their comments, we find a rather negative environment. Furthermore we show that the few users from the conspiracy echo chamber who interact with the debunking posts manifest a higher tendency to comment, in general. However, if we look at their commenting and liking rate – i.e., the daily number of comments and likes – we find that their activity in the conspiracy echo chamber increases after the interaction.

Thus, dissenting information online is ignored. Indeed, our results suggest that debunking information remains confined within the scientific echo chamber and that very few users of the conspiracy echo chamber interact with debunking posts. Moreover, the interaction seems to lead to an increasing interest in conspiracy-like content.

On our perspective the diffusion of bogus content is someway related to the increasing mistrust of people with respect to institutions, to the increasing level of functional illiteracy – i.e., the inability to understand information correctly – affecting western countries, as well as the combined effect of confirmation bias at work on a enormous basin of information where the quality is poor. According to these settings, current debunking campaigns as well as algorithmic solutions 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 with untrusted opponents in online discussion, the latter results in a major commitment with respect to their own echo chamber.

4.3 Materials and Methods

Ethics Statement

The entire data collection process is performed exclusively by means of the Facebook Graph API (109), which is publicly available and can be used through one's personal Facebook user account. We used only public available data (users with privacy restrictions are not included in our dataset). Data was downloaded from public Facebook pages that are public entities. Users' content contributing to such entities is also public unless the users' privacy settings specify otherwise and in that case it is not available to us. When allowed by users' privacy specifications, we accessed public personal information. However, in our study we used fully anonymized and aggregated data. We abided by the terms, conditions, and privacy policies of Facebook.

4.3.1 Data Collection

Data was downloaded from public Facebook pages that are accessible to anyone virtually. The entire data collection process is performed exclusively by means of the Facebook Graph API (109), which is publicly available and can be used through one's personal Facebook user account. The first category includes all pages diffusing conspiracy information – pages which disseminate controversial information, most often lacking supporting evidence and sometimes contradictory of the official news (i.e. conspiracy theories). The second category is that of scientific dissemination including scientific institutions and scientific press having the main mission to diffuse scientific knowledge. The third category contains all pages active in debunking false rumors online. We use this latter set as a testbed for the efficacy of debunking campaign. The exact

	Science	Conspiracy	Debunking	Total
Pages	83	330	66	479
Posts	262, 815	369, 420	50, 220	682, 455
Likes	463, 966, 540	145, 388, 131	4, 160, 674	613, 515, 345
Comments	22, 093, 692	8, 307, 643	488, 279	30, 889, 614
Likers	40, 466, 440	19, 386, 132	744, 023	52, 753, 883
Commenters	7, 223, 473	3, 166, 725	139, 168	9, 812, 332

Table 10: Breakdown of Facebook dataset. Number of pages, posts, likes, comments, likers, and commenters for science, conspiracy, and debunking pages.

breakdown of the data is presented in Table 10 (see Appendix B.2 for the complete list of pages).

4.3.2 Sentiment Classification

Data annotation consists in assigning some predefined labels to each data point. We selected a subset of 24,312 comments from the Facebook dataset (Table 10) and later used it to train a sentiment classifier. We used a user-friendly web and mobile devices annotation platform, Goldfinch[3] and engaged trustworthy English speakers, active on Facebook, for the annotations. The annotation task was to label each Facebook comment – isolated from its context – as *negative*, *neutral*, or *positive*. Each annotator had to estimate the emotional attitude of the user when posting a comment to Facebook. During the annotation process, the annotators performance was monitored in terms of the inter-annotator agreement and self-agreement, based on a subset of the comments which were intentionally duplicated. The annotation process resulted in 24,312 sentiment labeled comments, 6,555 of them annotated twice. We evaluate the self- and inter-annotator agreements in terms of Krippendorff's Alpha-reliability (133), which is a reliability coefficient able to measure the agreement of any number of annotators, often used in literature (134). $Alpha$ is defined

[3]provided by Sowa Labs: `http://www.sowalabs.com/`

as

$$Alpha = 1 - \frac{D_o}{D_e},$$

where D_o is the observed disagreement between annotators and D_e is the disagreement one would expect by chance. When annotators agree perfectly, $Alpha = 1$, and when the level of agreement equals the agreement by chance, $Alpha = 0$. In our case, $4,009$ comments were polled twice to two different annotators and are used to assess the inter-annotator agreement, for which $Alpha = 0.810$, while $2,546$ comments were polled twice to the same annotator and are used to asses the annotators' self-agreements, for which $Alpha = 0.916$.

We treat sentiment classification as an ordinal classification task with three ordered classes. We remind that ordinal classification is a form of multi-class classification where there is a natural ordering between the classes, but no meaningful numeric difference between them (110). We apply the wrapper approach, described in (114), with two linear-kernel Support Vector Machine (SVM) classifiers (98). SVM is a state-of-the-art supervised learning algorithm, well suited for large scale text categorization tasks, and robust on large feature spaces. The two SVM classifiers were trained to distinguish the extreme classes – *negative* and *positive* – from the rest – *neutral* plus *positive*, and *neutral* plus *negative*. During prediction, if both classifiers agree, they yield the common class, otherwise, if they disagree, the assigned class is *neutral*.

The sentiment classifier was trained and tuned on the training set of 19,450 annotated comments. The comments were processed into the standard Bag-of-Words (BoW) representation. The trained sentiment classifier was then evaluated on a disjoint test set of the remaining 4,862 comments. Three measures were used to evaluate the performance of the sentiment classifier:

1. The aforementioned $Alpha$
2. The $Accuracy$, defined as the fraction of correctly classified examples:

$$Accuracy = \frac{\langle -,- \rangle + \langle 0,0 \rangle + \langle +,+ \rangle}{N}$$

3. $\overline{F_1}(+,-)$, the macro-averaged F-score of the positive and negative classes, a standard evaluation measure (105) for sentiment classification tasks:
$$\overline{F_1}(+,-) = \frac{F_{1+} + F_{1-}}{2}$$
In general, F_1 is the harmonic mean of *Precision* and *Recall* for each class (111):
$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$
where *Precision* for class x is the fraction of correctly predicted examples out of all the predictions with class x:
$$Precision_x = \frac{\langle x, x \rangle}{\langle *, x \rangle}$$
and *Recall* for class x is the fraction of correctly predicted examples out of all the examples with actual class x:
$$Recall_x = \frac{\langle x, x \rangle}{\langle x, * \rangle}$$

The averaged evaluation are the followings: $Alpha = 0.589 \pm 0.017$, $Accuracy = 0.654 \pm 0.012$, and $\overline{F_1}(+,-) = 0.685 \pm 0.011$. The 95% confidence intervals are estimated from 10-fold cross validations.

4.3.3 Statistical Tools

Kaplan-Meier estimator. Let us define a random variable T on the interval $[0, \infty)$, indicating the time an event takes place. The cumulative distribution function (CDF), $F(t) = \mathbf{Pr}(T \leq t)$, indicates the probability that a subject selected at random will have a survival time less than or equal some stated value t. The survival function, defined as the complementary CDF (CCDF[4]) of T, is the probability of observing a survival time greater than some stated value t. To estimate this probability we use the *Kaplan–Meier estimator* (135). Let n_t denote the number of users at risk

[4]We remind that the CCDF of a random variable X is one minus the CDF, the function $f(x) = \mathbf{Pr}(X > x)$.

of stop commenting at time t, and let d_t denote the number of users that stop commenting precisely at t. Then, the conditional survival probability at time t is defined as $(n_t - d_t)/n_t$. Thus, if we have N observations at times $t_1 \leq t_2 \leq \cdots \leq t_N$, assuming that the events at times t_i are jointly independent, the Kaplan-Meier estimate of the survival function at time t is defined as

$$\hat{S}(t) = \prod_{t_i \leq t} (\frac{n_{t_i} - d_{t_i}}{n_{t_i}}),$$

with the convention that $\hat{S}(t) = 1, \quad \text{if} \;\; t < t_i$.

Comparison between power law distributions. Comparisons between power law distributions of two different quantities are usually carried out through log-likelihood ratio test (136) or Kolmogorov-Smirnov test (137). The former method relies on the ratio between the likelihood of a model fitted on the pooled quantities and the sum of the likelihoods of the models fitted on the two separate quantities, whereas the latter is based on the comparison between the cumulative distribution functions of the two quantities. However, both the afore-mentioned approaches take into account the overall distributions, whereas more often we are especially interested in the scaling parameter of the distribution, i.e. how the tail of the distribution behaves. Moreover, since the Kolmogorov-Smirnov test was conceived for continuous distributions, its application to discrete data gives biased p-values. For these reasons, in this paper we decide to compare our distributions by assess significant differences in the scaling parameters by means of a Wald test. The Wald test we conceive is defined as

$$H_0 : \hat{\alpha}_1 - \hat{\alpha}_2 = 0$$

$$H_1 : \hat{\alpha}_1 - \hat{\alpha}_2 \neq 0,$$

where $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are the estimates of the scaling parameters of the two powerlaw distributions. The Wald statistics,

$$\frac{(\hat{\alpha}_1 - \hat{\alpha}_2)^2}{VAR(\hat{\alpha}_1)},$$

	Likes	Comments	Shares
Power law	$\mathbf{-34,056.95}$	$\mathbf{-77,904.52}$	$\mathbf{-108,823.2}$
Poisson	$-22,143,084$	$-6,013,281$	$-109,045,636$
Lognormal	$-35,112.58$	$-82,619.08$	$-113,643.7$
Exponential	$-36,475.47$	$-87,859.85$	$-119,161.2$

Table 11: Goodness of fit for posts' attention patterns on conspiracy pages.

	Likes	Comments	Shares
Power law	$\mathbf{-33,371.53}$	$\mathbf{-2,537.418}$	$\mathbf{-4,994.981}$
Poisson	$-57,731,533$	$-497,016.2$	$-3,833,242$
Lognormal	$-34,016.76$	$-2,620.886$	$-5,126.515$
Exponential	$-35.330,76$	$-2,777.548$	$-5,415.722$

Table 12: Goodness of fit for posts' attention patterns on science pages.

where $VAR(\hat{\alpha}_1)$ is the variance of $\hat{\alpha}_1$, follows a χ^2 distribution with 1 degree of freedom. We reject the null hypothesis H_0 and conclude that there is a significant difference between the scaling parameters of the two distributions if the p-value of the Wald statistics is below a given significance level.

Attention Patterns. Different fits for the tail of the distributions have been taken into account (lognormal, Poisson, exponential, and power law). As for attention patterns related to posts, Goodness of fit tests based on the log-likelihood (137) have proved that the tails are best fitted by a power law distribution both for conspiracy and scientific news (see Tables 11 and 12).

Log-likelihoods of different attention patterns (likes, comments, and shares) are computed under competing distributions. The one with the higher log-likelihood is then the better fit (137). Log-likelihood ratio tests between power law and the other distributions yield positive ratios, and p-value computed using Vuong's method (138) are close to zero, indicating that the best fit provided by the power law distribution is not caused by statistical fluctuations. Lower bounds and scaling parameters have

	Likes		Comments		Shares	
	$\hat{x}_{min}$	$\hat{\alpha}$	$\hat{x}_{min}$	$\hat{\alpha}$	$\hat{x}_{min}$	$\hat{\alpha}$
Conspiracy	$8,995$	2.73	136	2.33	$1,800$	2.29
Science	$62,976$	2.78	$8,890$	3.27	$53,958$	3.41
t-stat	-	0.88	-	325.38	-	469.42
p-value	-	0.3477	-	$< 10^{-6}$	-	$< 10^{-6}$

Table 13: Power law fit of posts' attention patterns.

	Likes	Comments
Power law	$\mathbf{-24,044.40}$	$\mathbf{-57,274.31}$
Poisson	$-294,076.1$	$-334,825.6$
Lognormal	$-25,177.79$	$-62,415.91$
Exponential	$-28,068.09$	$-68,650.47$

Table 14: Goodness of fit for users' attention patterns on conspiracy pages.

been estimated via minimization of Kolmogorov-Smirnov statistics (137); the latter have been compared via Wald test (see Table 13).

As for users activity, Tables 14 and 15 list the fit parameters with various canonical distributions for both conspiracy and scientific news. Table 16 shows the power law fit parameters and summarizes the estimated lower bounds and scaling parameters for each distribution.

Cox-Hazard Model. The hazard function is modeled as

$$h(t) = h_0(t)\exp(\beta x),$$

where $h_0(t)$ is the baseline hazard and x is a dummy variable that takes value 1 when the user has been exposed to debunking and 0 otherwise. The hazards depend multiplicatively on the covariates, and $\exp(\beta)$ is the ratio of the hazards between users exposed and not exposed to debunking. The ratio of the hazards of any two users i and j is $\exp(\beta(x_i - x_j))$, and is called the *hazard ratio*. This ratio is assumed to be constant over time, hence the name of proportional hazard. When we consider exposure to debunking by means of likes, the estimated β is 0.72742 $(s.e. =$

	Likes	**Comments**
Power law	$\mathbf{-222,763.1}$	$\mathbf{-42,901.23}$
Poisson	$-5,027,337$	$-260,162.7$
Lognormal	$-231,319.1$	$-46,752.34$
Exponential	$-249,771.4$	$-51,345.45$

Table 15: Goodness of fit for users' attention patterns on science pages.

	Likes		**Comments**	
	$\hat{x}_{min}$	$\hat{\alpha}$	$\hat{x}_{min}$	$\hat{\alpha}$
Conspiracy	900	4.07	45	2.93
Science	900	3.25	45	3.07
t-stat		952.56		17.89
p-value		$< 10^{-6}$		2.34×10^{-5}

Table 16: Power law fit of users' attention patterns.

$0.01991,\ p < 10^{-6}$) and the corresponding hazard ratio, $\exp(\beta)$, between users exposed and not exposed is 2.07, indicating that users not exposed to debunking are 2.07 times more likely to stop consuming conspiracy news. Goodness of fit for the Cox Proportional Hazard Model has been assessed by means of Likelihood ratio test, Wald test, and Score test which provided p-values close to zero. Fig 19 *(left)* shows the fit of the Cox proportional hazard model when the lifetime is computed on likes.

Moreover, if we consider exposure to debunking by means of comments, the estimated β is 0.56748 ($s.e. = 0.02711,\ p < 10^{-6}$) and the corresponding hazard ratio, $\exp(\beta)$, between users exposed and not exposed is 1.76, indicating that users not exposed to debunking are 1.76 times more likely to stop consuming conspiracy news. Goodness of fit for the Cox Proportional Hazard Model has been assessed by means of Likelihood ratio test, Wald test, and Score test, which provided p-values close to zero. Fig 19 *(right)* shows the fit of the Cox proportional hazard model when the lifetime is computed on comments.

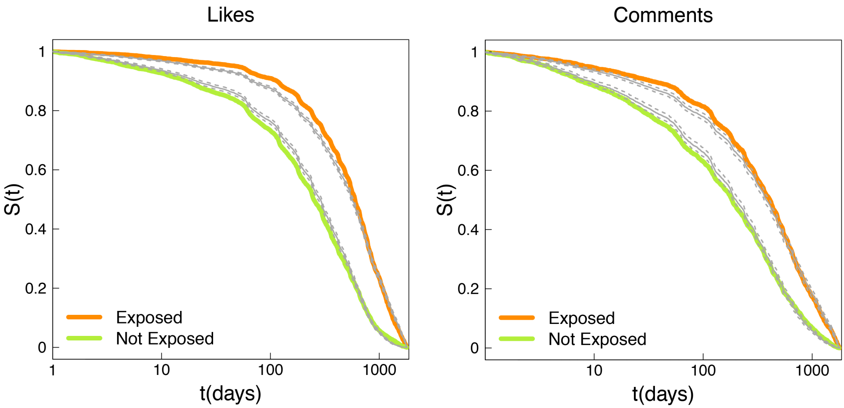


Figure 19: Cox-Hazard Model Kaplan-Meier estimates of survival functions of users who interacted *(exposed, orange)* and did not *(not exposed, green)* with debunking and fits of the Cox proportional hazard model. Persistence of users is computed both on likes *(left)* and comments *(right)*.

CHAPTER
FIVE

CONCLUSIONS AND FUTURE WORKS

In this chapter we discuss all the follow-ups of the research line presented in the previous sections. We know that users tend to form polarized groups of like-minded people (73). Immersed in these echo chambers, they frame and reinforce their world view (16; 75), acquire information confirming their preferred narrative (69), and ignore dissenting information (36). Moreover, debating with like-minded people have been shown to negatively influence their emotions and to burst group polarization (74).

At this stage of the research we have a good understanding of the social dynamics behind group polarization and of the proliferation of unsubstantiated rumors online. However, the problem still remains complex and intricate. Currently we are striving to determine key metrics able to identify echo chambers just accounting for users' interaction with posts and how specific topics are perceived as critical by different groups of users. This effort aims at framing information campaigns based on a quantitatively tailored narrative.

To explore the criticality of topics, we may measure how different echo chambers debate around the same topics. In the latest years one of the most controversial subject is anthropogenic climate change. The issue is politically polarizing (139), especially in nations with organized denial

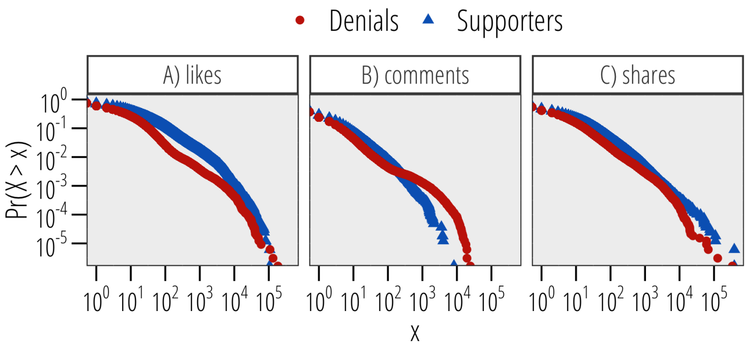


Figure 20: Attention Patterns. Complementary cumulative distribution functions (CCDFs) of the number of likes, comments, and shares received by posts belonging to Denials *((red)* and Supporters *(blue)* pages.

campaigns (140). Studies have examined these dynamics in the blogosphere (141), online comments (142), and Twitter (143), but not Facebook. Therefore we want to extend the analysis on climate change polarization to Facebook[1]. We focus on echo chambers emerging from interactions with the pages of 76 blogs supporting/promoting climate science and 69 pages of blogs denying/questioning climate change and science (see Appendix B.3 for the complete list of pages) over a time span of six years, a total of 500K posts with more than 4M users liking and commenting.

In Fig. 20 we show how posts of the two narratives get consumed in terms of number of likes, comments and shares. The plots capture similar interaction patterns, all distributions are heavy tailed. Claims supporting and denying climate change reverberate in a comparable way and receive a similar volume of attention on the Facebook platform.

We now want to characterize users' mobility across the different types of content. Hence, for each user we count the total number of her likes and comments on posts supporting and denying climate science and de-

[1]This material is based on an about to be submitted co-authored paper. Authors: Fabiana Zollo, Alessandro Bessi, Michela Del Vicario, Antonio Scala, Guido Caldarelli, Walter Quattrociocchi and Riley Dunlap. FZ conceived and designed the experiments; performed the analysis and interpreted the results; contributed to writing the manuscript.

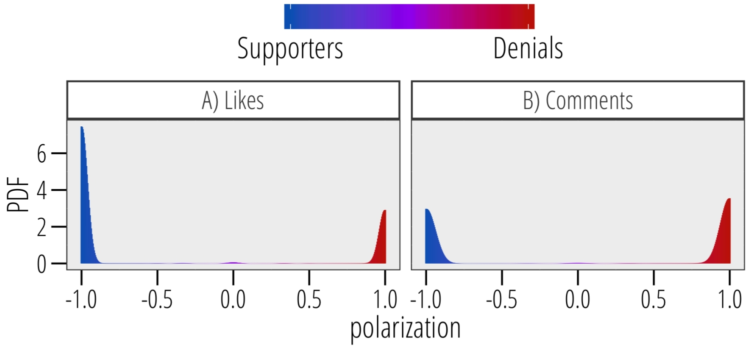


Figure 21: Probability Density Functions (PDFs) of the polarization of Supporters and Denials computed both on likes *(left)* and on comments *(right)*.

fine the user polarization ρ as the ratio of likes on a specific category with respect to the user's total number of likes. The more the value is close to 1 (or -1) the more the user interacts with one of the two narratives. Fig. 21 shows the Probability Density Function of users' activity on the two categories. We notice that distributions are sharply bimodal. Users focus on one kind of narrative and ignore the other one. Information belonging to the two narratives gets absorbed by different and isolated groups.

The tendency of users to focus on their favorite narratives and form polarized groups might alter the way in which certain topics are absorbed. To validate this aspect, we analyze how the subject of a post is presented to the users. We make use of IBM AlchemyAPI (144) to extract semantic metadata from posts content. In particular we extract the sentiment and main concepts discussed within each post of the dataset whether it has a textual description or a link to an external document.

Fig. 22 shows the sentiment distribution of posts on both Supporters and Denials pages. We may observe a slightly negative pattern for both categories, although more pronounced for Denials rather than for Supporters. Notice that we are referring to the way in which subjects are discussed into the post; we are not taking into account the sentiment that the post may elicit in the reader, or the sentiment of users involved in the

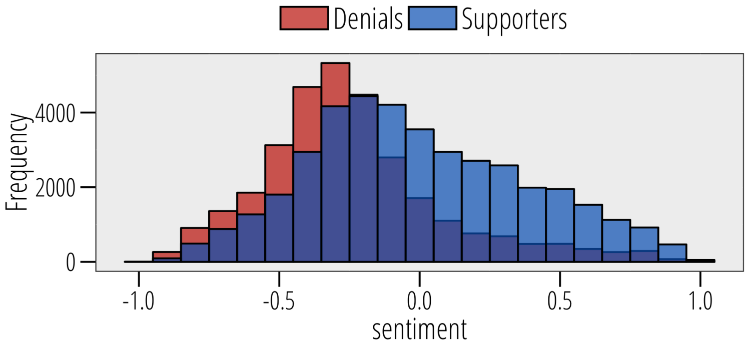


Figure 22: Sentiment distribution of posts on Denials and Supporters pages. The sentiment is defined in the range $[-1, 1]$, where -1 is negative, 0 is neutral, and 1 is positive.

discussion.

To make explicit how the Supporters and Denials echo chambers perceive the different debated issues, we now focus on concepts discussed by pages of both narratives. Fig. 23 shows all the concepts shared by posts of both echo chambers. For each concept we compute its average sentiment i.e., the mean of the sentiment of all the posts where it appears. Concepts are ordered by taking into account the difference between the average sentiment on Supporters posts and that on Denials ones. Thus, concepts at the bottom are discussed in a much more similar way on both categories, differently from the top.

Our findings show that the discussion is polarized and both echo chambers express different emotions with respect to the same topics. In particular, the technique introduced in Fig. 23 could be of great interest to identify the most controversial topics. Indeed, it is thoroughly likely that the greater the emotional distance between the same concept in two echo chambers, the greater the polarization of users involved in the discussion. Therefore, this distance may become a key marker to locate crucial topics and understand how to deal with them.

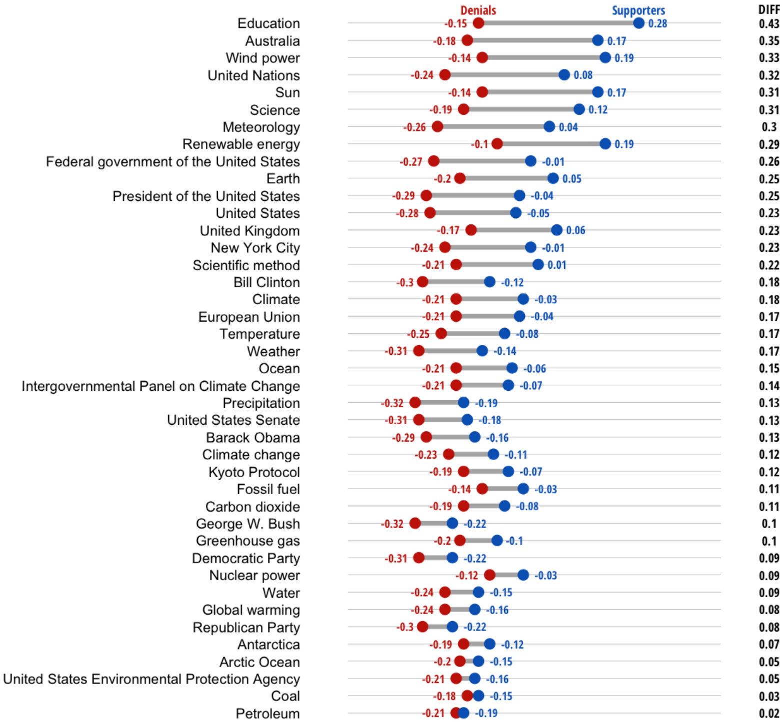


Figure 23: Concepts shared by Supporters and Denials echo chambers. For each concept we show its average sentiment (computed over all the posts where it appears). Concepts are ordered by taking into account the difference between the average sentiment on Supporters posts and that on Denials ones.

APPENDIX

A

OTHER WORKS

A.1 The Spreading of Misinformation Online

This section is based on a co-authored paper published on the Proceedings of the National Academy of Sciences of the United States of America (PNAS) (73)[1].

A.1.1 Main Results and Discussion

The wide availability of user-provided content in online social media facilitates the aggregation of people around common interests, world-views, and narratives. However, the World Wide Web is a fruitful environment for the massive diffusion of unverified rumors. In this work, using a massive quantitative analysis of Facebook, we show that information related to distinct narratives generates homogeneous and polarized communities having similar information consumption patterns.

In particular, we focus on how Facebook users consume information related to two distinct narratives: scientific and conspiracy news. We find that, although consumers of scientific and conspiracy stories present similar consumption patterns with respect to content, cascade dynamics

[1]FZ contributed to design and perform research; provide analytic tools; analyze data; write the paper. See (73) for further details.

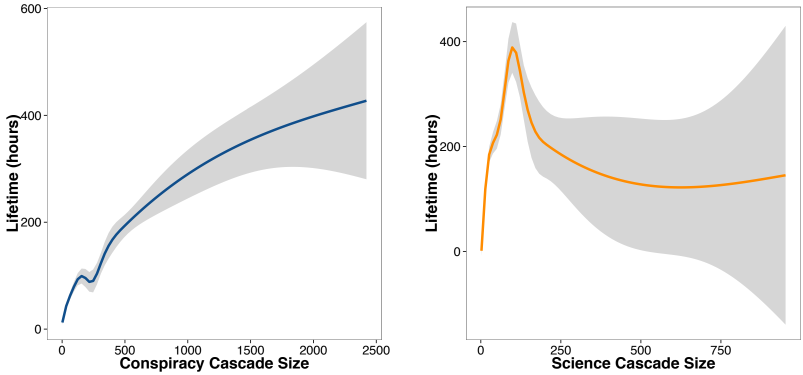


Figure 24: Lifetime as a function of the cascade size for conspiracy news *(left)* and science news *(right)*. Science news quickly reaches a higher diffusion; a longer lifetime does not correspond to a higher level of interest. Conspiracy rumors are assimilated more slowly and show a positive relation between lifetime and size.

differ. Selective exposure to content is the primary driver of content diffusion and generates the formation of *echo chambers*. Indeed, homogeneity appears to be the primary driver for the diffusion of contents and each echo chamber has its own cascade dynamics.

Fig. 24 shows the lifetime as a function of the cascade size. 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. These results suggest that news assimilation differs according to the categories. Science news is usually assimilated, i.e., it reaches a higher level of diffusion quickly, and a longer lifetime does not correspond to a higher level of interest. Conversely, conspiracy rumors are assimilated more slowly and show a positive relation between lifetime and size. For both science and conspiracy news, we compute the size as a function of the lifetime and confirm that differentiation in the

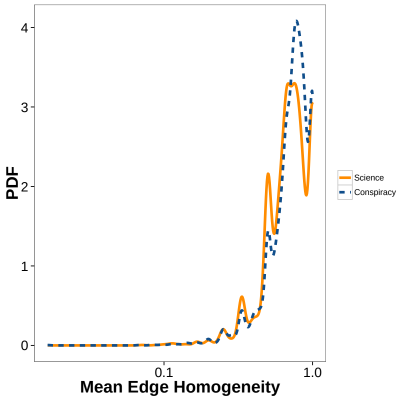


Figure 25: PDF of edge homogeneity for science *(orange)* and conspiracy *(blue)* news. Homogeneity paths are dominant on the whole cascades for both scientific and conspiracy news.

sharing patterns is content-driven, and that for conspiracy there is a positive relation between size and lifetime.

We next examine the social determinants that drive sharing patterns and we focus on the role of homo- geneity in friendship networks. Fig. 25 shows the PDF of the mean-edge homogeneity, computed for all cascades of science news and conspiracy theories. It shows that the majority of links between consecutively sharing users is homogeneous. 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 nonhomogeneous links is favorable toward the former ones. However, the probability of close to zero mean-edge homogeneity is quite small. Contents tend to circulate only inside the echo chamber.

Hence, to further characterize the role of homogeneity in shaping sharing cascades, we compute cascade size as a function of mean-edge homogeneity for both science and conspiracy news (Fig. 26). In science

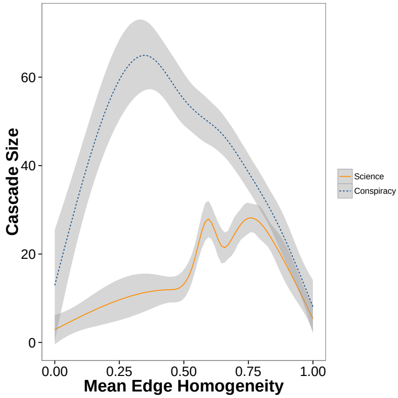


Figure 26: Cascade size as a function of edge homogeneity for science *(orange)* and conspiracy *(blue)* news.

news, higher levels of mean-edge homogeneity in the interval $(0.5, 0.8)$ correspond to larger cascades, but in conspiracy 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.

Our findings show that users mostly tend to select and share content according to a specific narrative and to ignore the rest. This suggests that the determinant for the formation of echo chambers is confirmation bias. To model this mechanism we now introduce a percolation model of rumor spreading to account for homogeneity and polarization and we show that homogeneity and polarization are the main determinants for predicting cascades' size.

A.2 Users Polarization on Facebook and YouTube

This section is based on a co-authored paper published on PLoS ONE (145)[2].

A.2.1 Main Results and Discussion

Users online tend to select information that support and adhere their beliefs, and to form polarized groups sharing the same view – e.g. echo chambers. Algorithms for content promotion may favour this phenomenon, by accounting for users preferences and thus limiting the exposure to unsolicited contents. To shade light on this question, we perform a comparative study on how same contents (videos) are consumed on different online social media – i.e. Facebook and YouTube – over a sample of 12M of users.

We focus on Facebook posts linking Youtube videos reported on Science and Conspiracy pages. We then compare the users interaction with these videos on both platforms. Fig. 27 shows the Probability Density Functions (PDFs) of about 12M users and on how they distribute their comments on Science and Conspiracy posts (polarization) on both Facebook and YouTube. We observe sharply peaked bimodal distributions. Users concentrate their activity on one of the two narratives. Indeed the percentage of polarized users is equal to 93.6% on Facebook and 87.8% on YouTube; therefore, two well separated communities support competing narratives in both online social networks. Content has a polarizing effect: users focus on specific types of content and aggregate in separated groups – echo chambers – independently of the platform and content promotion algorithm.

We extend our analysis by investigating the polarization dynamics – i.e., how users become polarized comment after comment. On both platforms, we observe that some users interact only with a specific kind

[2]FZ contributed materials and analysis tools; write the paper. See (145) for further details.

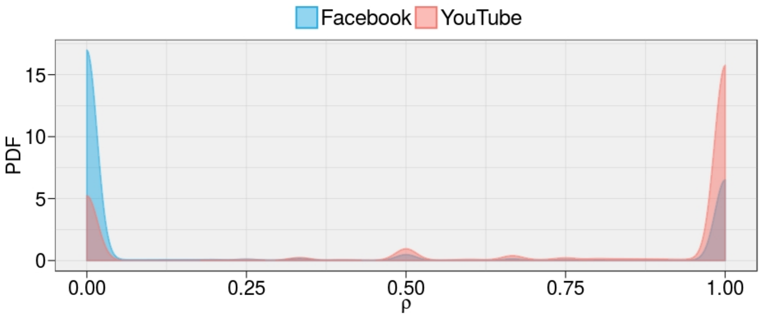


Figure 27: Polarization on Facebook and YouTube. The PDFs of the polarization ρ show that the vast majority of users is polarized towards one of the two conflicting narratives – i.e. Science and Conspiracy – on both Facebook and YouTube.

of content since the beginning, whereas others start their commenting activity by switching between contents supporting different narratives. The vast majority of the latter – after the initial switching phase – starts consuming mainly one type of information, becoming polarized towards one of the two conflicting narratives.

Our findings show that content drives the emergence of echo chambers on both platforms. Moreover, we show that the users' commenting patterns are accurate predictors for the formation of echo chambers.

A.3 Homophily and Polarization in the Age of Misinformation

This section is based on a co-authored paper to appear on the European Physical Journal Special Topics (EPJ ST) in 2016 (146)[3]. Preliminary results were also published in the Proceedings of the 24th International Conference on World Wide Web (72).

A.3.1 Main Results and Discussion

The World Economic Forum listed massive digital misinformation as one of the main threats for our society. The spreading of unsubstantiated rumors may have serious consequences on public opinion such as in the case of rumors about Ebola causing disruption to health-care workers. In this work we target Facebook to characterize information consumption patterns of 1.2M Italian users with respect to verified (science news) and unverified (conspiracy news) contents. Through a thorough quantitative analysis we provide important insights about the anatomy of the system across which misinformation might spread. In particular, we show that users' engagement on verified (or unverified) content correlates with the number of friends having similar consumption patterns i.e., homophily.

Fig. 28 shows the linear relationship between the fraction of friends polarized on the same category of the user and the logarithm of her activity. Thus, we check whether for a polarized user the fraction of polarized friends in her category can be predicted by means of a linear regression model where the explanatory variable is a logarithmic transformation of the number of likes θ i.e., $y = \beta_0 + \beta_1 \log(\theta)$. Coefficients are estimated using ordinary least squares and they are –with the corresponding standard errors inside the round brackets– $\hat{\beta}_0 = 0.70\ (0.005)$ and $\hat{\beta}_1 = 0.043\ (0.001)$, with $R^2 = 0.95$, for users polarized towards science, and $\hat{\beta}_0 = 0.71\ (0.003)$ and $\hat{\beta}_1 = 0.047\ (0.0006)$, with $R^2 = 0.98$, for users polarized towards conspiracy.

[3]FZ contributed materials and analysis tools; wrote the paper. See (72; 146) for further details.

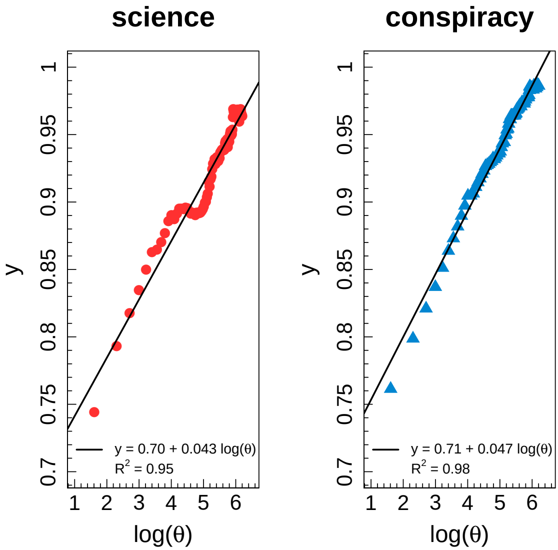


Figure 28: Predicting the fraction of friends of users polarized on science (left) and on conspiracy (right).

Finally, we measure how this social system responded to the injection of 4, 709 false information. Fig. 29 illustrates the average value of the polarization, $avg(\rho)$, for increasing levels of shares; more precisely, we compute the average polarization of all the users who liked troll posts with a number of shares greater than x. We find an increasing trend that starts from an average polarization of ~ 0.6 and asymptotically stabilizes at about ~ 0.73; the average polarization starts to increase sharply at $x \sim 20$ and saturates at $x \sim 200$. Users exposed to conspiracy stories seem to be more prone to diffuse intentionally false information.

Our findings show that the frequent (and selective) exposure to specific kind of content (polarization) is a good proxy for the detection of homophile clusters where certain kind of rumors are more likely to spread.

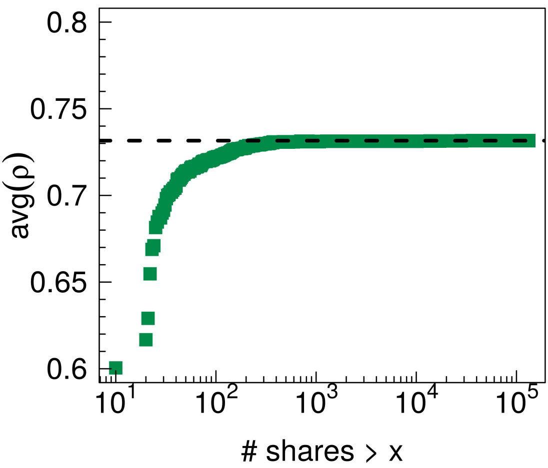


Figure 29: Average polarization of users who liked troll posts (intentionally false information). Notice that the polarization increases with the number of shares, indicating that very popular posts containing false information are mostly supported by conspiracy users.

A.4 Trend of Narratives in the Age of Misinformation

This section is based on a co-authored paper published on PLoS ONE (75)[4].

A.4.1 Main Results and Discussion

Social media enabled a direct path from producer to consumer of contents changing the way users get informed, debate, and shape their worldviews. Such a *disintermediation* might weaken consensus on social relevant issues in favor of rumors, mistrust, or conspiracy thinking – e.g., chem-trails inducing global warming, the link between vaccines and autism, or the New World Order conspiracy. Previous studies pointed out that consumers of conspiracy-like content are likely to aggregate in homophile clusters – i.e., echo-chambers.

Along this path we study, by means of a thorough quantitative analysis, how different topics are consumed inside the conspiracy echo-chamber in the Italian Facebook. Through a semi-automatic topic extraction strategy, we show that the most consumed contents semantically refer to four specific categories: *environment*, *diet*, *health*, and *geopolitics*. Fig. 30 shows the backbone of the co-occurrence term network, where different colors indicate nodes belonging to different conspiracy categories. We find similar consumption patterns by comparing users activity (likes and comments) on posts belonging to these different semantic categories.

Finally, we analyze the relationship between the engagement of a user – i.e. the number of likes she left on conspiracy posts – and how her activity is distributed across categories. We model users mobility across the distinct topics finding that the more a user is active, the more he is likely to span on all categories. Fig. 31 shows that the more a conspiracy user is engaged the more his activity spread on the overall corpus. Indeed, once inside a conspiracy narrative users tend to embrace the overall corpus.

[4]FZ contributed to perform the experiments, analyze data, provide materials and analysis tools, write the paper. See (75) for further details.

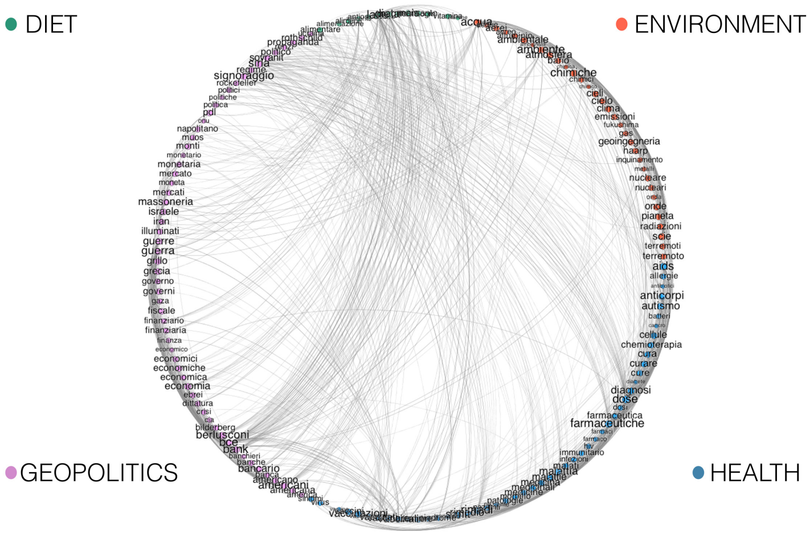


Figure 30: Backbone of conspiracy terms co-occurence network. Different colors indicate nodes belonging to different semantic category according to the output of the supervised tagging. In particular, purple nodes belong to geopolitics, red nodes to environment, blue nodes to health, and green to diet.

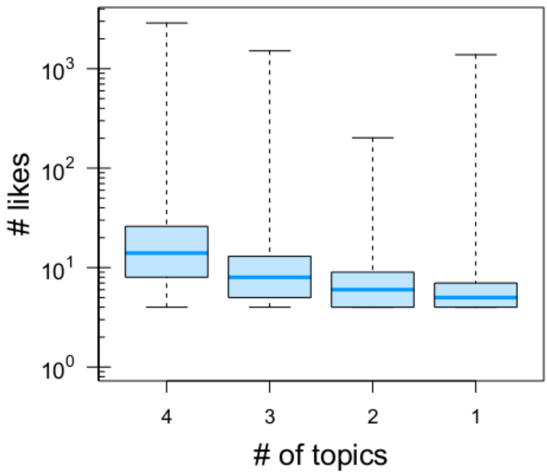


Figure 31: Engagement and mobility across semantic categories. Light blue lines represent the median of the likes distributions; pale blue shaded boxes represent the interquartile range (25–75 percentile); horizontal bars represent the extremes of the distributions. Users active on four categories are $15,510$; users active on three categories are $20,929$; users active on two categories are $21,631$; and users active on one category are $9,980$.

APPENDIX

B

DATASETS

B.1 Science & Conspiracy on the Italian Facebook

We provide the full list of Facebook pages of our Italian dataset. Table 17 lists scientific pages, while Table 18 lists conspiracy pages.

Table 17: Scientific news sources of the Italian dataset.

	Page Name	Facebook ID
1	Scientificast.it	129133110517884
2	CICAP	32775139194
3	OggiScienza	106965734432
4	Query	128523133833337
5	Gravità Zero	138484279514358
6	COELUM Astronomia	81631306737
7	MedBunker	246240278737917
8	In Difesa della Sperimentazione Animale	365212740272738
9	Italia Unita per la Scienza	492924810790346
10	Scienza Live	227175397415634
11	La scienza come non l'avete mai vista	230542647135219
12	LIBERASCIENZA	301266998787
13	Scienze Naturali	134760945225
14	Perché vaccino	338627506257240
15	Le Scienze	146489812096483
16	Vera scienza	389493082245
17	Scienza in rete	84645527341
18	Galileo, giornale di scienza e problemi globali	94897729756
19	Scie Chimiche: Informazione Corretta	351626174626
20	Complottismo? No grazie	399888818975
21	INFN - Istituto Nazionale di Fisica Nucleare	45086217578
22	Signoraggio: informazione corretta	279217954594

23	JFK informazione corretta	113204388784459
24	Scetticamente	146529622080908
25	Vivisezione e Sperimentazione Animale, verità e menzogne	548684548518541
26	Medici Senza Frontiere	65737832194
27	Task Force Pandora	273189619499850
28	VaccinarSI	148150648573922
29	Lega Nerd	165086498710
30	Super Quark	47601641660
31	Curiosità Scientifiche	595492993822831
32	Minerva - Associazione di Divulgazione Scientifica	161460900714958
33	Pro-Test Italia	221292424664911
34	Uniti per la Ricerca	132734716745038

Table 18: Conspiracy news sources of the Italian dataset.

	Page Name	Facebook ID
1	Scienza di Confine	188189217954979
2	CSSC - Cieli Senza Scie Chimiche	253520844711659
3	STOP ALLE SCIE CHIMICHE	199277020680
4	Vaccini Basta	233426770069342
5	Tanker Enemy	444154468988487
6	SCIE CHIMICHE	68091825232
7	MES Dittatore Europeo	194120424046954
8	Lo sai	126393880733870
9	AmbienteBio	109383485816534
10	Eco(R)esistenza	203737476337348
11	curarsialnaturale	159590407439801
12	La Resistenza	256612957830788
13	Radical Bio	124489267724876
14	Fuori da Matrix	123944574364433
15	Graviola Italia	130541730433071
16	Signoraggio.it	278440415537619
17	Informare Per Resistere	101748583911
18	Sul Nuovo Ordine Mondiale	340262489362734
19	Avvistamenti e Contatti	352513104826417
20	Umani in Divenire	195235103879949
21	Nikola Tesla - il SEGRETO	108255081924
22	Teletrasporto	100774912863
23	PNL e Ipnosi	150500394993159
24	HAARP - controllo climatico	117166361628599
25	Sezione Aurea, Studio di Energia Vibrazionale	113640815379825
26	PER UNA NUOVA MEDICINA	113933508706361
27	PSICOALIMENTARSI E CURARSI NATURALMENTE	119866258041409
28	La nostra ignoranza la LORO forza.	520400687983468
29	HIV non causa AIDS	121365461259470
30	Sapere un Dovere	444729718909881
31	V per Verità	223425924337104
32	Genitori veg	211328765641743
33	Operatori di luce	195636673927835
34	Coscienza Nuova	292747470828855
35	Aprite Gli Occhi	145389958854351
36	Neovitruvian	128660840526907
37	CoscienzaSveglia	158362357555710
38	Medicinenon	248246118546060
39	TERRA REAL TIME	208776375809817

B.2 Science & Conspiracy on the US Facebook

We provide the full list of Facebook pages of our US dataset. Table 19 lists conspiracy pages, while Table 20 lists scientific pages, and Table 21 lists debunking pages.

Table 19: Conspiracy news sources of the US dataset.

	Page Name	**Facebook ID**
1	Spirit Science and Metaphysics	171274739679432
2	Spirit Science	210238862349944
3	The Conspiracy Archives	262849270399655
4	iReleaseEndorphins	297719273575542
5	World of Lucid Dreaming	98584674825
6	The Science of Spirit	345684712212932
7	Esoteric Philosophy	141347145919527
8	9/11 Truth Movement	259930617384687
9	Great Health The Natural Way	177320665694370
10	New World Order News	111156025645268
11	Freedom Isn't Free on FB	634692139880441
12	Skeptic Society	224391964369022
13	The Spiritualist	197053767098051
14	Anonymous World Wide	494931210527903
15	The Life Beyond Earth	152806824765696
16	Illuminati Exposed	298088266957281
17	Illuminating Souls	38466722555
18	Alternative Way	119695318182956
19	Paranormal Conspiracies	455572884515474
20	CANNABIS CURES CANCERS!	115759665126597
21	Natural Cures Not Medicine	1104995126306864
22	CTA Conspiracy Theorists' Association	515416211855967
23	Illuminati Killers	478715722175123
24	Conspiracy 2012 & Beyond	116676015097888
25	GMO Dangers	182443691771352
26	The Truthers Awareness	576279865724651
27	Exposing the truth about America	385979414829070
28	Occupy Bilderberg	231170273608124
29	Speak the Revolution	422518854486140
30	I Don't Trust The Government	380911408658563
31	Sky Watch Map	417198734990619
32	| truthaholics	201546203216539
33	UFO Phenomenon	419069998168962
34	Conspiracy Theories & The Illuminati	117611941738491
35	Lets Change The World	625843777452057
36	Makaveli The Prince Killuminati	827000284010733
37	It's A New Day	116492031738006
38	New world outlawz - killuminati soldiers	422048874529740
39	The Government's bullshit. Your argument is invalid.	173884216111509
40	America Awakened	620954014584248
41	The truth behold	466578896732948
42	Alien Ufo And News	334372653327841
43	Anti-Bilderberg Resistance Movement	161284443959494
44	The Truth Unleashed	431558836898020
45	Anti GMO Foods and Fluoride Water	366658260094302
46	STOP Controlling Nature	168168276654316
47	9/11 Blogger	109918092364301

48	9/11 Studies and Outreach Club at ASU	507983502576368
49	9/11 Truth News	120603014657906
50	Abolish the FDA	198124706875206
51	AboveTopSecret.com	141621602544762
52	Activist Post	128407570539436
53	Alliance for Natural Health USA	243777274534
54	All Natural & Organic. Say No To Toxic Chemicals.	323383287739269
55	Alternative Medicine	219403238093061
56	Alternative World News Network	154779684564904
57	AltHealthWORKS	318639724882355
58	American Academy of Environmental Medicine	61115567111
59	American Association of Naturopathic Physicians	14848224715
60	Ancient Alien Theory	147986808591048
61	Ancient Aliens	100140296694563
62	Ancient Astronaut Theory	73808938369
63	The Anti-Media	156720204453023
64	Anti Sodium Fluoride Movement	143932698972116
65	Architects & Engineers for 9/11 Truth	59185411268
66	Association of Accredited Naturopathic Medical Colleges (AANMC)	60708531146
67	Autism Media Channel	129733027101435
68	Babes Against Biotech	327002374043204
69	Bawell Alkaline Water Ionizer Health Benefits	447465781968559
70	CancerTruth	348939748204
71	Chemtrails Awareness	12282631069
72	Collective Evolution	131929868907
73	Conspiracy Theory With Jesse Ventura	122021024620821
74	The Daily Sheeple	114637491995485
75	Dr. Bronner's Magic Soaps	33699882778
76	Dr. Joseph Mercola	114205065589
77	Dr. Ronald Hoffman	110231295707464
78	Earth. We are one.	149658285050501
79	Educate Inspire Change	467083626712253
80	Energise for Life: The Alkaline Diet Experts!	99263884780
81	Exposing The Truth	175868780941
82	The Farmacy	482134055140366
83	Fluoride Action Network	109230302473419
84	Food Babe	132535093447877
85	Global Research (Centre for Research on Globalization)	200870816591393
86	GMO Inside	478981558808326
87	GMO Just Say No	1390244744536466
88	GreenMedInfo.com	111877548489
89	Healthy Holistic Living	134953239880777
90	I Fucking Love Truth	445723122122920
91	InfoWars	80256732576
92	Institute for Responsible Technology	355853721234
93	I Want To Be 100% Organic	431825520263804
94	Knowledge of Today	307551552600363
95	La Healthy Living	251131238330504
96	March Against Monsanto	566004240084767
97	Millions Against Monsanto by OrganicConsumers.org	289934516904
98	The Mind Unleashed	432632306793920
99	Moms Across America	111116155721597
100	Moms for Clean Air/Stop Jet Aerosol Spraying	1550135768532988
101	Natural Society	191822234195749
102	Non-GMO Project	55972693514
103	Occupy Corporatism	227213404014035
104	The Open Mind	782036978473504
105	Organic Consumers Association	13341879933
106	Organic Health	637019016358534

107	The Organic Prepper	435427356522981
108	PreventDisease.com	199701427498
109	Raw For Beauty	280583218719915
110	REALfarmacy.com	457765807639814
111	ReThink911	581078305246370
112	Sacred Geometry and Ancient Knowledge	363116270489862
113	Stop OC Smart Meters	164620026961366
114	The Top Information Post	505941169465529
115	The Truth About Vaccines	133579170019140
116	Truth Teller	278837732170258
117	Veterans Today	170917822620
118	What Doctors Don't Tell You	157620297591924
119	Wheat Belly	209766919069873
120	Why don't you try this?	202719226544269
121	WND	119984188013847
122	WorldTruth.TV	114896831960040
123	Zeitgeist	32985985640
124	Ancient Origins	530869733620642
125	Astrology Answers	413145432131383
126	Astrology News Service	196416677051124
127	Autism Action Network	162315170489749
128	Awakening America	406363186091465
129	Awakening People	204136819599624
130	Cannabinoids Cure Diseases & The Endocannabinoid System Makes It Possible.	322971327723145
131	Celestial Healing Wellness Center	123165847709982
132	Chico Sky Watch	149772398420200
133	A Conscious awakening	539906446080416
134	Conspiracy Syndrome	138267619575029
135	Conspiracy Theory: Truth Hidden in Plain Sight, and Army of SATAN	124113537743088
136	Cosmic Intelligence-Agency	164324963624932
137	C4ST	371347602949295
138	Deepak Chopra	184133190664
139	Dr. Mehmet Oz	35541499994
140	Earth Patriot	373323356902
141	Electromagnetic Radiation Safety	465980443450930
142	EMF Safety Network	199793306742863
143	End Time Headlines	135010313189665
144	Young Living Essential Oils	29796911981
145	Exposing Bilderberg 2012	300498383360728
146	Exposing The Illuminati	196087297165394
147	Exposing Satanic World Government	529736240478567
148	FEMA Camps Exposed	285257418255898
149	Fight Against Illuminati And New World Order	195559810501401
150	FitLife.tv	148518475178805
151	GMO Free USA	402058139834655
152	Holistic Health	105497186147476
153	The Illuminati	543854275628660
154	Illuminati Mind Control	499866223357022
155	Intelwars	130166550361356
156	Natural Solutions Foundation	234136166735798
157	NWO Truth Radio	135090269995781
158	Occupy Bilderberg 2012	227692450670795
159	Operation: Awakening- The Global Revolution	287772794657070
160	The Paradigm Shift	221341527884801
161	PositiveMed	177648308949017
162	Press TV	145097112198751
163	The Resistance	394604877344757
164	Rima E. Laibow, M.D. - Save My Life Dr. Rima	107527312740569
165	RT America	137767151365

166	Ruble's Wonderings - Forbidden Archeology & Science	265422293590870
167	Seekers Of Truth	736499966368634
168	Spiritual Ecology	261982733906722
169	Spiritualer.com	531950866874307
170	Take Back Your Power	269179579827247
171	There is a cure for Cancer, but it is not FDA approved. Phoenix Tears work!	395190597537
172	True Activist	129370207168068
173	Truth Exposed Radio	173823575962481
174	Truth Movement	161389033958012
175	Truth Network	271701606246002
176	Wake up call	276404442375280
177	We Should Ban GMOs	516524895097781
178	vactruth.com	287991907988
179	Veterans Today Truth Warriors	645478795537771
180	4 Foot Farm Blueprint	1377091479178258
181	Dawning Golden Crystal Age	127815003927694
182	Occupy Your Mind	393849780700637
183	We do not Forgive. We do not Forget. We are Anonymous. Expect Us.	134030470016833
184	Health Impact News	469121526459635
185	NaturalNews.com	35590531315
186	World for 9/11 Truth	38411749990
187	Beware of Disinformation	558882824140805
188	Citizens For Legitimate Government	93486533659
189	Cureyourowncancer.org	535679936458252
190	Juicing Vegetables	172567162798498
191	Quantum Prophecies	323520924404870
192	AIM Integrative Medicine	137141869763519
193	Autism Nutrition Research Center	1508552969368252
194	The Canary Party	220071664686886
195	Chemtrail Research	247681531931261
196	Chemtrail Watchers	77065926441
197	Children's Medical Safety Research Institute	790296257666848
198	Contaminated Vaccines	686182981422650
199	Dane Wigington	680418385353616
200	David Icke	147823328841
201	David Icke Books Limited	191364871070270
202	David Icke - Headlines	1421025651509652
203	Disinformation Directory	258624097663749
204	The Drs. Wolfson	1428115297409777
205	Educate, Inspire & Change. The Truth Is Out There, Just Open Your Eyes	111415972358133
206	Focus for Health Foundation	456051981200997
207	Generation Rescue	162566388038
208	Geoengineering Watch	448281071877305
209	Global Skywatch	128141750715760
210	The Greater Good	145865008809119
211	The Health Freedom Express	450411098403289
212	Homegrown Health	190048467776279
213	Intellihub	439119036166643
214	The Liberty Beacon	222092971257181
215	International Medical Council on Vaccination	121591387888250
216	International Medical Council on Vaccination - Maine Chapter	149150225097217
217	Medical Jane	156904131109730
218	Mississippi Parents for Vaccine Rights	141170989357307
219	My parents didn't put me in time-out, they whooped my ass!	275738084532
220	National Vaccine Information Center	143745137930
221	The Raw Feed Live	441287025913792
222	Rinf.com	154434341237962
223	SANEVAX	139881632707155
224	Things pro-vaxers say	770620782980490

225	Unvaccinated America	384030984975351
226	Vaccine Injury Law Project	295977950440133
227	Vermont Coalition for Vaccine Choice	380959335251497
228	9/11: The BIGGEST LIE	129496843915554
229	Agent Orange Activists	644062532320637
230	Age of Autism	183383325034032
231	AutismOne	199957646696501
232	Awakened Citizen	481936318539426
233	Best Chinese Medicines	153901834710826
234	Black Salve	224002417695782
235	Bought Movie	144198595771434
236	Children Of Vietnam Veterans Health Alliance	222449644516926
237	Collective-Evolution Shift	277160669144420
238	Doctors Are Dangerous	292077004229528
239	Dr. Tenpenny on Vaccines	171964245890
240	Dr Wakefield's work must continue	84956903164
241	EndoRIOT	168746323267370
242	Enenews	126572280756448
243	Expanded Consciousness	372843136091545
244	Exposing the truths of the Illuminati II	157896884221277
245	Family Health Freedom Network	157276081149274
246	Fearless Parent	327609184049041
247	Food Integrity Now	336641393949
248	Four Winds 10	233310423466959
249	Fukushima Explosion What You Do Not Know	1448402432051510
250	The Golden Secrets	250112083847
251	Health Without Medicine & Food Without Chemicals	304937512905083
252	Higher Perspective	488353241197000
253	livingmaxwell	109584749954
254	JFK Truth	1426437510917392
255	New World Order Library | NWO Library	194994541179
256	No Fluoride	117837414684
257	Open Minds Magazine	139382669461984
258	Organic Seed Alliance	111220277149
259	Organic Seed Growers and Trade Association	124679267607065
260	RadChick Radiation Research & Mitigation	260610960640885
261	The REAL Institute - Max Bliss	328240720622120
262	Realities Watch	647751428644641
263	StormCloudsGathering	152920038142341
264	Tenpenny Integrative Medical Centers (TIMC)	144578885593545
265	Vaccine Epidemic	190754844273581
266	VaccineImpact	783513531728629
267	Weston A. Price Foundation	58956225915
268	What On Earth Is Happening	735263086566914
269	The World According to Monsanto	70550557294
270	Truth Theory	175719755481
271	Csglobe	403588786403016
272	Free Energy Truth	192446108025
273	Smart Meter Education Network	630418936987737
274	The Mountain Astrologer magazine	112278112664
275	Alberta Chemtrail Crusaders	1453419071541217
276	Alkaline Us	430099307105773
277	Americas Freedom Fighters	568982666502934
278	Anti-Masonic Party Founded 1828	610426282420191
279	Cannabidiol OIL	241449942632203
280	Cancer Compass˜An Alternate Route	464410856902927
281	Collective Evolution Lifestyle	1412660665693795
282	Conscious Life News	148270801883880
283	Disclosure Project	112617022158085

284	Dr. Russell Blaylock, MD	123113281055091
285	Dumbing Down People into Sheeple	123846131099156
286	Expand Your Consciousness	351484988331613
287	Fluoride: Poison on Tap	1391282847818928
288	Gaiam TV	182073298490036
289	Gary Null & Associates	141821219197583
290	Genesis II Church of Health & Healing (Official)	115744595234934
291	Genetic Crimes Unit	286464338091839
292	Global Healing Center	49262013645
293	Gluten Free Society	156656676820
294	GMO Free Oregon	352284908147199
295	GMO Journal	113999915313056
296	GMO OMG	525732617477488
297	GreenMedTV	1441106586124552
298	Healing The Symptoms Known As Autism	475607685847989
299	Health Conspiracy Radio	225749987558859
300	Health and Happiness	463582507091863
301	Jesse Ventura	138233432870955
302	Jim Humble	252310611483446
303	Kid Against Chemo	742946279111241
304	Kids Right To Know Club	622586431101931
305	The Master Mineral Solution of the 3rd Millennium	527697750598681
306	Millions Against Monsanto Maui	278949835538988
307	Millions Against Monsanto World Food Day 2011	116087401827626
308	Newsmax Health	139852149523097
309	Non GMO journal	303024523153829
310	Nurses Against ALL Vaccines	751472191586573
311	Oath Keepers	182483688451972
312	Oath Keepers of America	1476304325928788
313	The Organic & Non-GMO Report	98397470347
314	Oregon Coast Holographic Skies Informants	185456364957528
315	Paranormal Research Project	1408287352721685
316	Politically incorrect America	340862132747401
317	(Pure Energy Systems) PES Network, Inc.	183247495049420
318	Save Hawaii from Monsanto	486359274757546
319	Sayer Ji	205672406261058
320	SecretSpaceProgram	126070004103888
321	SPM Southern Patriots MIlitia	284567008366903
322	Thrive	204987926185574
323	Truth Connections	717024228355607
324	Truth Frequency	396012345346
325	Truthstream Media.com	193175867500745
326	VT Right To Know GMOs	259010264170581
327	We Are Change	86518833689
328	Wisdom Tribe 7 Walking in Wisdom.	625899837467523
329	World Association for Vaccine Education	1485654141655627
330	X Tribune	1516605761946273

Table 20: Scientific news sources of the US dataset.

	Page Name	Facebook ID
1	AAAS - The American Association for the Advancement of Science	19192438096
2	AAAS Dialogue on Science, Ethics and Religion	183292605082365
3	Armed with Science	228662449288
4	AsapSCIENCE	162558843875154
5	Bridge to Science	185160951530768
6	EurekAlert!	178218971326
7	Food Science	165396023578703
8	Food Science and Nutrition	117931493622
9	I fucking love science	367116489976035
10	LiveScience	30478646760
11	Medical Laboratory Science	122670427760880
12	National Geographic Magazine	72996268335
13	National Science Foundation (NSF)	30037047899
14	Nature	6115848166
15	Nature Education	109424643283
16	Nature Reviews	328116510545096
17	News from Science	100864590107
18	Popular Science	60342206410
19	RealClearScience	122453341144402
20	Science	96191425588
21	Science and Mathematics	149102251852371
22	Science Channel	14391502916
23	Science Friday	10862798402
24	Science News Magazine	35695491869
25	Science-Based Medicine	354768227983392
26	Science-fact	167184886633926
27	Science, Critical Thinking and Skepticism	274760745963769
28	Science: The Magic of Reality	253023781481792
29	ScienceDaily	60510727180
30	ScienceDump	111815475513565
31	ScienceInsider	160971773939586
32	Scientific American magazine	22297920245
33	Scientific Reports	143076299093134
34	Sense About Science	182689751780179
35	Skeptical Science	317015763334
36	The Beauty of Science & Reality.	215021375271374
37	The Flame Challenge	299969013403575
38	The New York Times - Science	105307012882667
39	Wired Science	6607338526
40	All Science, All the Time	247817072005099
41	Life's Little Mysteries	373856446287
42	Reason Magazine	17548474116
43	Nature News and Comment	139267936143724
44	Astronomy Magazine	108218329601
45	CERN	169005736520113
46	Citizen Science	200725956684695
47	Cosmos	143870639031920
48	Discover Magazine	9045517075
49	Discovery News	107124643386
50	Genetics and Genomics	459858430718215
51	Genetic Research Group	193134710731208
52	Medical Daily	189874081082249
53	MIT Technology Review	17043549797
54	NASA - National Aeronautics and Space Administration	54971236771
55	New Scientist	235877164588
56	Science Babe	492861780850602

57	ScienceBlogs	256321580087
58	Science, History, Exploration	174143646109353
59	Science News for Students	136673493023607
60	The Skeptics Society & Skeptic Magazine	23479859352
61	Compound Interest	1426695400897512
62	Kevin M. Folta	712124122199236
63	Southern Fried Science	411969035092
64	ThatsNonsense.com	107149055980624
65	Science & Reason	159797170698491
66	ScienceAlert	7557552517
67	Discovery	6002238585
68	Critical Thinker Academy	175658485789832
69	Critical Thinking and Logic Courses in US Core Public School Curriculum	171842589538247
70	Cultural Cognition Project	287319338042474
71	Foundation for Critical Thinking	56761578230
72	Immunization Action Coalition	456742707709399
73	James Randi Educational Foundation	340406508527
74	NCSE: The National Center for Science Education	185362080579
75	Neil deGrasse Tyson	7720276612
76	Science, Mother Fucker. Science	228620660672248
77	The Immunization Partnership	218891728752
78	Farm Babe	1491945694421203
79	Phys.org	47849178041
80	Technology Org	218038858333420
81	Biology Fortified, Inc.	179017932138240
82	The Annenberg Public Policy Center of the University of Pennsylvania	123413357705549
83	Best Food Facts	200562936624790

Table 21: Debunking news sources of the US dataset.

	Page Name	**Facebook ID**
1	Refutations to Anti-Vaccine Memes	414643305272351
2	Boycott Organic	1415898565330025
3	Contrails and Chemtrails:The truth behind the myth	391450627601206
4	Contrail Science	339553572770902
5	Contrail Science and Facts - Stop the Fear Campaign	344100572354341
6	Debunking Denialism	321539551292979
7	The Farmer's Daughter	350270581699871
8	GMO Answers	477352609019085
9	The Hawaii Farmer's Daughter	660617173949316
10	People for factual GMO truths (pro-GMO)	255945427857439
11	The Questionist	415335941857289
12	Scientific skepticism	570668942967053
13	The Skeptic's Dictionary	195265446870
14	Stop the Anti-Science Movement	1402181230021857
15	The Thinking Person's Guide to Autism	119870308054305
16	Antiviral	326412844183079
17	Center for Inquiry	5945034772
18	The Committee for Skeptical Inquiry	50659619036
19	Doubtful News	283777734966177
20	Hoax-Slayer	69502133435
21	I fucking hate pseudoscience	163735987107605
22	The Genetic Literacy Project	126936247426054
23	Making Sense of Fluoride	549091551795860
24	Metabunk	178975622126946
25	Point of Inquiry	32152655601
26	Quackwatch	220319368131898

27	Rationalwiki	226614404019306
28	Science-Based Pharmacy	141250142707983
29	Skeptical Inquirer	55675557620
30	Skeptic North	141205274247
31	The Skeptics' Guide to the Universe	16599501604
32	Society for Science-Based Medicine	552269441534959
33	Things anti-vaxers say	656716804343725
34	This Week in Pseudoscience	485501288225656
35	Violent metaphors	537355189645145
36	wafflesatnoon.com	155026824528163
37	We Love GMOs and Vaccines	1380693538867364
38	California Immunization Coalition	273110136291
39	Exposing PseudoAstronomy	218172464933868
40	CSICOP	157877444419
41	The Panic Virus	102263206510736
42	The Quackometer	331993286821644
43	Phil Plait	251070648641
44	Science For The Open Minded	274363899399265
45	Skeptic's Toolbox	142131352492158
46	Vaccine Nation	1453445781556645
47	Vaximom	340286212731675
48	Voices for Vaccines	279714615481820
49	Big Organic	652647568145937
50	Chemtrails are NOT real, idiots are.	235745389878867
51	Sluts for Monsanto	326598190839084
52	Stop Homeopathy Plus	182042075247396
53	They Blinded Me with Pseudoscience	791793554212187
54	Pro-Vaccine Shills for Big Pharma, the Illumanati, Reptilians, and the NWO	709431502441281
55	Pilots explain Contrails - and the Chemtrail Hoax	367930929968504
56	The Skeptical Beard	325381847652490
57	The Alliance For Food and Farming	401665083177817
58	Skeptical Raptor	522616064482036
59	Anti-Anti-Vaccine Campaign	334891353257708
60	Informed Citizens Against Vaccination Misinformation	144023769075631
61	Museum of Scientifically Proven Supernatural and Paranormal Phenomena	221030544679341
62	Emergent	375919272559739
63	Green State TV	128813933807183
64	Kavin Senapathy	1488134174787224
65	vactruth.com Exposed	1526700274269631
66	snopes.com	241061082705085

B.3 Climate Change on Facebook

We provide the full list of pages of our Facebook dataset about climate change. Table 22 lists pages of denying/questioning climate change and science, while Table 23 lists pages supporting/promoting climate science.

Table 22: Pages supporting anthropogenic global warming.

	Page Name	Facebook ID
1	Climate Central	57984115023
2	Climate Change For Dummies	253660441493203
3	Climate Change Guide	209071469166691
4	Denial101x	671468936283032
5	Global Warming Fact of the Day	239542442866980
6	I Heart Climate Scientists	332593866775047
7	IPCC Intergovernmental Panel on Climate Change	165091923539860
8	NASA Climate Change	353034908075
9	NPS Climate Change Response	527593347362646
10	RealClearScience	122453341144402
11	ScienceAlert	7557552517
12	Skeptical Science	317015763334
13	The Skeptics' Guide to the Universe	16599501604
14	Society of Environmental Journalists	433086126732239
15	World Meteorological Organization	71741701887
16	American Meteorological Society	74679966771
17	ARC Centre of Excellence for Climate System Science	197303993749021
18	Ask a Climate Scientist	282058025204543
19	AtmosNews - NCAR UCAR - Atmospheric & Earth System Science	78148787037
20	Climate Change Facts	87694967110
21	Climate Change and the Pacific Islands	975669532452905
22	Climate Change Policy & Practice	192140072102
23	The Climate Denial Crock of the Week	270080745702
24	Climate Reality	153278754738777
25	Climate Speakers Network	579252922169321
26	COP21	1514890778722700
27	European Association of Environmental and Resource Economists EAERE	113938735290092
28	Global Sustainability Foundation	303269799822796
29	I Heart Climate Scientists	332593866775047
30	InsideClimate News	245371732167183
31	International Institute for Environment and Development IIED	111963037878
32	International Institute for Sustainable Development IISD	72656094247
33	International Organization for Migration - Micronesia	138751482902381
34	Monash Simple Climate Model	1213399748677520
35	NCAR Computational and Information Systems Laboratory	1491699071092530
36	NCSE: The National Center for Science Education	185362080579
37	Northwest Climate Science Center	1417878408476380
38	PAST Lab	170667903076996
39	School of Integrated Climate System Sciences SICSS	121603291211738
40	ScienceAlert	7557552517
41	South-Central Climate Science Center	237652062962951
42	Southern Climate Impacts Planning Program	119105441532461
43	Southern Fried Science	411969035092
44	Springer Climate	278122429054474
45	Sustainable Development Policy & Practice / Post-2015 Development Agenda	126385310775420

46	UCARConnect	362743340521245
47	Uwapei Global Warming Campaign	1614051862144030
48	World Climate Research Programme	222477407818862
49	Agenda21.it	235726426493102
50	Bureau of Meteorology	170992086298033
51	Center for Climate Change Communication	359055448297
52	CleanTechnica	154039257947286
53	Citizens' Climate Lobby 3rd Coast Region	388244997921404
54	Climate Frontlines	206138405843
55	Climate Progress	187005858017014
56	CSIRO	142468583842
57	Environmental Science & Policy at Taylor & Francis	161789877175634
58	Five-Feet.org	464387463731327
59	Global Warming Climate Change Report	96279696758
60	The GLOBE Program	233891673345693
61	Katharine Hayhoe	1463215773903300
62	Michael E. Mann	221222081267335
63	National Climate Assessment	271900656322465
64	National Oceanic and Atmospheric Administration NOAA	201357451715
65	National Renewable Energy Laboratory	73775159896
66	National Science Foundation NSF	30037047899
67	Nature Climate Change	133045436728938
68	NOAA Climate.Gov	320631784698200
69	NOAA Great Lakes Environmental Research Laboratory	154479014611347
70	NOAA Libraries	132665076873047
71	NOAA National Centers for Environmental Information - Climate	348738721824427
72	NOAA Satellite and Information Service	226849284022023
73	NOAA Science On a Sphere	252976835082
74	U.S. Global Change Research Program	143460545677845
75	USGS News: Climate Change	206340569511295
76	What to Do about Climate Change	155811667893723

Table 23: Pages denying anthropogenic global warming.

	Page Name	**Facebook ID**
1	American Thinker	144317282271701
2	Australian Climate Madness	117005835006924
3	Breitbart	95475020353
4	Carbon Dioxide	64162630683
5	Carbon Tax! Whats Next?	154174501315763
6	Center for Industrial Progress	215077831880321
7	CFACT	140379955280
8	Climate Change Fraud	322696661147
9	Climate Change Hoax	169840503192972
10	Climate change is natural	313827422002524
11	Climate Change LIES	152483204848827
12	Climategate	226309168331
13	ClimateRealists.com	54260521733
14	Climate Depot	149314838564139
15	Climate-Gate	186103788219
16	Climate Hustle	1088950287806110
17	Climate News	306212519483530
18	Conservative Tribune	519305544814653
19	Cornwall Alliance for the Stewardship of Creation	12326763660
20	The Daily Caller	182919686769
21	David Icke	147823328841
22	Expose Agenda 21	239480352758933

23	The Federalist	157843634416312
24	The Federalist Papers	107705785934333
25	Friends of Science	244675788944611
26	The Galileo Movement	101728306584541
27	Global Warming a Scam?	10150111201315300
28	Global Warming Climate Change Hoaxers	942367752456567
29	Heartland Institute	16775672689
30	Human Events	212436860202
31	InfoWars	80256732576
32	The John Birch Society	196291400410701
33	Knowledge Replaces Fear and Ignorance	409494392454244
34	Left Exposed	634007473381128
35	Moonbattery	150964594926416
36	NaturalNews.com	35590531315
37	The New American Magazine	146909368666979
38	NGP-Next Gen Patriots	248737065182352
39	No Frakking Consensus	112150972159981
40	Oil Sands Action	590117937780756
41	Principia Scientific International	223487287746649
42	Renewables Versus Fossil Fuels Facts and Myths	1561787114111510
43	The Revolution	142868065759441
44	Steven Crowder	15139936162
45	We Know The Secrets of The Federal Reserve	178303555571465
46	wattsupwiththat Anthony Watts	133662869999306
47	WND formerly WorldNetDaily	119984188013847
48	World Around Us - Victoria	1612317212341320
49	You Might be a Conservative	131401483600716
50	100 Percent FED Up	311190048935167
51	4timesayear	360297500785985
52	CO2 is not a pollutant	140149506100162
53	Agenda 21 Exposed	150975061627200
54	AgEnders NJ - Agenda 21 Enders of New Jersey	126990307392585
55	AgEnders OK - Agenda 21 Enders of Oklahoma	296069487166798
56	Agenders Tennessee	213562718663812
57	Axed: The End of Green	488280944530354
58	Blue Beats Green	231287880367215
59	Climate change.the conspiracy theory.	249511709222
60	Climate Hustle	1088950287806110
61	Crying Wolf Documentary	209426629082
62	The Liberty Beacon	222092971257181
63	Lord Christopher Monckton 3rd Viscount Monckton of Brenchley	167541331466
64	Moms Against Agenda 21	243717865657674
65	Ohio Agenda 21 Watch	181003108643120
66	Stop Agenda 21 in TN	457433564301109
67	Stop Agenda 21- Washington State	253570204742793
68	Stop UN Agenda 21! Stop ICLEI!	284021125057
69	Climate of Corruption: Politics and Power Behind the Global Warming Hoax	201761106537309

REFERENCES

[1] E. Katz and P. F. Lazarsfeld, *Personal Influence, The part played by people in the flow of mass communications*. Transaction Publishers, 1970.

[2] P. F. Lazarsfeld, B. Berelson, and H. Gaudet, "The peoples choice: how the voter makes up his mind in a presidential campaign.," *New York Columbia University Press*, 1968.

[3] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," *Journal of consumer research*, vol. 34, no. 4, pp. 441–458, 2007.

[4] S. Buckingham Shum, K. Aberer, A. Schmidt, S. Bishop, P. Lukowicz, S. Anderson, Y. Charalabidis, J. Domingue, S. de Freitas, I. Dunwell, B. Edmonds, F. Grey, M. Haklay, M. Jelasity, A. Karpiataenko, J. Kohlhammer, J. Lewis, J. Pitt, R. Sumner, and D. Helbing, "Towards a global participatory platform," *The European Physical Journal Special Topics*, vol. 214, no. 1, pp. 109–152, 2012.

[5] P. Levy, *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Perseus Publishing, 1999.

[6] T. W. Malone and M. Klein, "Harnessing collective intelligence to address global climate change," *Innovations: Technology, Governance, Globalization*, vol. 2:3, pp. 15–26, 2007.

[7] N. Shadbolt, W. Hall, J. A. Hendler, and W. H. Dutton, "Web science: a new frontier," *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 371, March 2013.

[8] J. H. Kuklinski, P. J. Quirk, J. Jerit, D. Schwieder, and R. F. Rich, "Misinformation and the currency of democratic citizenship," *Journal of Politics*, vol. 62, no. 3, pp. 790–816, 2000.

[9] W. L. Howell, "Digital wildfires in a hyperconnected world," Tech. Rep. Global Risks 2013, World Economic Forum, 2013.

[10] B. Nyhan and J. Reifler, "When corrections fail: The persistence of political misperceptions," *Political Behavior*, vol. 32, no. 2, pp. 303–330, 2010.

[11] J. P. Moore, "The dangers of denying HIV," *Nature*, vol. 459, no. 7244, pp. 168–168, 2009.

[12] L. M. Bogart and S. Thorburn, "Are HIV/AIDS conspiracy beliefs a barrier to HIV prevention among african americans?," *JAIDS Journal of Acquired Immune Deficiency Syndromes*, vol. 38, no. 2, pp. 213–218, 2005.

[13] S. Kalichman, "Denying AIDS," *Conspiracy theories, pseudoscience, and human tragedy*, pp. 8–76, 2009.

[14] W. H. O. M. Centre, "Ebola: Experimental therapies and rumoured remedies." Situation Assessment, August 2014.

[15] L. Lessig, *Code: And other laws of cyberspace*. ReadHowYouWant. com, 2009.

[16] C. R. Sunstein, "The law of group polarization," *Journal of political philosophy*, vol. 10, no. 2, pp. 175–195, 2002.

[17] R. Spears, M. Lea, and S. Lee, "De-individuation and group polarization in computer-mediated communication," *British Journal of Social Psychology*, vol. 29, no. 2, pp. 121–134, 1990.

[18] E. Aronson, *The social animal*. Macmillan, 2003.

[19] L. Ross and R. E. Nisbett, *The person and the situation: Perspectives of social psychology*. Pinter & Martin Publishers, 2011.

[20] A. Caplin and J. Leahy, "Miracle on sixth avenue: information externalities and search," *The Economic Journal*, vol. 108, no. 446, pp. 60–74, 1998.

[21] S. E. Asch, "Opinions and social pressure," *Readings about the social animal*, vol. 193, pp. 17–26, 1955.

[22] D. M. Kahan, "Social influence, social meaning, and deterrence," *Virginia Law Review*, pp. 349–395, 1997.

[23] G. A. Akerlof, J. L. Yellen, and M. L. Katz, "An analysis of out-of-wedlock childbearing in the united states," *The Quarterly Journal of Economics*, pp. 277–317, 1996.

[24] J. A. Stoner, "Risky and cautious shifts in group decisions: The influence of widely held values," *Journal of Experimental Social Psychology*, vol. 4, no. 4, pp. 442–459, 1968.

[25] J. Giles, "Making the links," *Nature*, vol. 488, August 2012.

[26] D. Lazer, A. Pentland, L. Adamic, A. Sinan, B. Albert-László, B. Devon, C. Nicholas, C. Noshir, F. James, G. Myron, J. Tony, K. Gary, M. Michael, R. Deb, and V. A. Marshall, "Computational social science," *Science*, vol. 323, pp. 721–723, February 2009.

[27] M. Savage and R. Burrows, "The coming crisis of empirical sociology," *Sociology*, vol. 41, January 2007.

[28] M. Huberty, "Awaiting the second big data revolution." BRIE Working Paper, May 2014.

[29] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," *Mckinsey & Company*, 2011.

[30] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec, "Can cascades be predicted?," in *Proceedings of the 23rd international conference on World Wide Web*, pp. 925–936, International World Wide Web Conferences Steering Committee, 2014.

[31] P. A. Dow, L. A. Adamic, and A. Friggeri, "The anatomy of large Facebook cascades.," in *ICWSM*, 2013.

[32] A. Friggeri, L. A. Adamic, D. Eckles, and J. Cheng, "Rumor cascades," in *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, Association for the Advancement of Artificial Intelligence, 2014.

[33] S. Goel, A. Anderson, J. Hofman, and D. Watts, "The structural virality of online diffusion," *Preprint*, vol. 22, p. 26, 2013.

[34] "Snopes.com," October 2015.

[35] "Emergent.info," October 2015.

[36] F. Zollo, A. Bessi, M. Del Vicario, A. Scala, G. Caldarelli, L. Shekhtman, S. Havlin, and W. Quattrociocchi, "Debunking in a world of tribes," *arXiv preprint arXiv:1510.04267*, 2015.

[37] S. Aral, L. Muchnik, and A. Sundararajan, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21544–21549, 2009.

[38] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, pp. 415–444, 2001.

[39] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg, "Structural diversity in social contagion," *Proceedings of the National Academy of Sciences*, vol. 109, no. 16, pp. 5962–5966, 2012.

[40] E. Adar, L. Zhang, L. A. Adamic, and R. M. Lukose, "Implicit structure and the dynamics of blogspace," in *Workshop on the weblogging ecosystem*, vol. 13, pp. 16989–16995, 2004.

[41] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: quantifying influence on Twitter," in *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 65–74, ACM, 2011.

[42] D. Centola, "The spread of behavior in an online social network experiment," *Science*, vol. 329, no. 5996, pp. 1194–1197, 2010.

[43] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, "Emotional contagion.," *Current Directions in Psychological Science*, 1993.

[44] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, pp. 8788–8790, June 2014.

[45] M. Salathé and S. Khandelwal, "Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control," *PLoS Comput Biol*, vol. 7, no. 10, p. e1002199, 2011.

[46] J. H. Fowler and N. A. Christakis, "Cooperative behavior cascades in human social networks," *Proceedings of the National Academy of Sciences*, vol. 107, no. 12, pp. 5334–5338, 2010.

[47] W. Quattrociocchi, G. Caldarelli, and A. Scala, "Opinion dynamics on interacting networks: media competition and social influence," *Scientific reports*, vol. 4, 2014.

[48] M. J. Salganik, P. S. Dodds, and D. J. Watts, "Experimental study of inequality and unpredictability in an artificial cultural market," *Science*, vol. 311, no. 5762, pp. 854–856, 2006.

[49] S. Aral and D. Walker, "Identifying influential and susceptible members of social networks," *Science*, vol. 337, no. 6092, pp. 337–341, 2012.

[50] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, no. 7415, pp. 295–298, 2012.

[51] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in *Proceedings of the 13th international conference on World Wide Web*, pp. 491–501, ACM, 2004.

[52] J. Leskovec, M. McGlohon, C. Faloutsos, N. S. Glance, and M. Hurst, "Patterns of cascading behavior in large blog graphs.," in *SDM*, vol. 7, pp. 551–556, SIAM, 2007.

[53] B. Golub and M. O. Jackson, "Using selection bias to explain the observed structure of internet diffusions," *Proceedings of the National Academy of Sciences*, vol. 107, no. 24, pp. 10833–10836, 2010.

[54] D. Liben-Nowell and J. Kleinberg, "Tracing information flow on a global scale using internet chain-letter data," *Proceedings of the National Academy of Sciences*, vol. 105, no. 12, pp. 4633–4638, 2008.

[55] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on Digg and Twitter social networks.," *ICWSM*, vol. 10, pp. 90–97, 2010.

[56] S. Goel, D. J. Watts, and D. G. Goldstein, "The structure of online diffusion networks," in *Proceedings of the 13th ACM conference on electronic commerce*, pp. 623–638, ACM, 2012.

[57] D. J. Watts, *Everything Is Obvious: How Common Sense Fails Us*. Random House LLC, 2012.

[58] L. Adamic *et al.*, "The diffusion of support in an online social movement: Evidence from the adoption of equal-sign profile pictures," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1741–1750, ACM, 2015.

[59] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in *Proceedings of the 21st international conference on World Wide Web*, pp. 519–528, ACM, 2012.

[60] C. R. Sunstein, *Republic.com 2.0*. Princeton University Press, 2009.

[61] E. Pariser, *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

[62] E. Bakshy, S. Messing, and L. Adamic, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, p. aaa1160, 2015.

[63] S. Knobloch-Westerwick, "Selective exposure and reinforcement of attitudes and partisanship before a presidential election," *Journal of Communication*, vol. 62, no. 4, pp. 628–642, 2012.

[64] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises.," *Review of general psychology*, vol. 2, no. 2, p. 175, 1998.

[65] X. L. Dong, E. Gabrilovich, K. Murphy, V. Dang, W. Horn, C. Lugaresi, S. Sun, and W. Zhang, "Knowledge-based trust: Estimating the trustworthiness of web sources," *Proceedings of the VLDB Endowment*, vol. 8, no. 9, pp. 938–949, 2015.

[66] O. Erich and W. Udi, "News feed fyi: Showing fewer hoaxes," January 2015.

[67] B. Nyhan, J. Reifler, S. Richey, and G. L. Freed, "Effective messages in vaccine promotion: a randomized trial," *Pediatrics*, vol. 133, no. 4, pp. e835–e842, 2014.

[68] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi, "Collective attention in the age of (mis)information," *Computers in Human Behavior*, 2015.

[69] A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Science vs conspiracy: Collective narratives in the age of misinformation," *PLoS ONE*, vol. 10, 02 2015.

[70] A. Bessi, G. Caldarelli, M. Del Vicario, A. Scala, and W. Quattrociocchi, "Social determinants of content selection in the age of (mis)information," *Proceedings of SOCINFO 2014*, 2014.

[71] A. Bessi, A. Scala, L. Rossi, Q. Zhang, and W. Quattrociocchi, "The economy of attention in the age of (mis) information," *Journal of Trust Management*, vol. 1, no. 1, pp. 1–13, 2014.

[72] A. Bessi, F. Petroni, M. Del Vicario, F. Zollo, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Viral misinformation: The role of homophily and polarization," in *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pp. 355–356, 2015.

[73] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, "The spreading of misinformation online," *Proceedings of the National Academy of Sciences*, vol. 113, no. 3, pp. 554–559, 2016.

[74] F. Zollo, P. K. Novak, M. Del Vicario, A. Bessi, I. Mozetič, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Emotional dynamics in the age of misinformation," *PLoS ONE*, vol. 10, 09 2015.

[75] A. Bessi, F. Zollo, M. Del Vicario, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Trend of narratives in the age of misinformation," *PLoS One*, 2015.

[76] B. D. Loader, A. Vromen, and M. A. Xenos, "The networked young citizen: social media, political participation and civic engagement," *Information, Communication & Society*, vol. 17, no. 2, pp. 143–150, 2014.

[77] M. J. Magro, "A review of social media use in e-government," *Administrative Sciences*, vol. 2, no. 2, pp. 148–161, 2012.

[78] S. V. Scott and W. J. Orlikowski, "Entanglements in practice: Performing anonymity through social media," *Management Information Systems Research Center, Carlson School of Management, University of Minnesota*, 2014.

[79] T. Brabazon, "Digital distinctiveness," in *Unique Urbanity?*, pp. 85–92, Springer, 2015.

[80] E. Farley, F. Grady, D. S. Miller, R. O'Connor, H. Schneider, M. Spikes, C. Constantinou, *et al.*, "What happens when everyone yields the power of information? - Handouts," *School of Journalism, Stony Brook University*, 2014.

[81] S. Meraz, "Is there an élite hold? Traditional media to social media agenda setting influence in blog networks," *Journal of Computer-Mediated Communication*, vol. 14, no. 3, pp. 682–707, 2009.

[82] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 US election: divided they blog," in *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43, ACM, 2005.

[83] E. Colleoni, A. Rozza, and A. Arvidsson, "Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data," *Journal of Communication*, vol. 64, no. 2, pp. 317–332, 2014.

[84] A. Jøsang, W. Quattrociocchi, and D. Karabeg, "Taste and trust," in *Trust Management V*, pp. 312–322, Springer, 2011.

[85] W. Quattrociocchi, R. Conte, and E. Lodi, "Opinions manipulation: Media, power and gossip," *Advances in Complex Systems*, vol. 14, no. 04, pp. 567–586, 2011.

[86] W. Quattrociocchi, R. Conte, and E. Lodi, "Simulating opinion dynamics in heterogeneous communication systems," *arXiv preprint arXiv:1101.3085*, 2011.

[87] S. Yardi and D. Boyd, "Dynamic debates: An analysis of group polarization over time on Twitter," *Bulletin of Science, Technology & Society*, vol. 30, no. 5, pp. 316–327, 2010.

[88] E. Bonabeau, "Decisions 2.0: The power of collective intelligence," *MIT Sloan management review*, vol. 50, no. 2, pp. 45–52, 2009.

[89] J. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few*. Abacus, 2005.

[90] "Why Operation Jade Helm 15 is freaking out the Internet – and why it shouldn't be." The Washington Post, November 2015.

[91] J. Byford, *Conspiracy Theories: A Critical Introduction*. Palgrave Macmillan, 2011.

[92] G. A. Fine, V. Campion-Vincent, and C. Heath, *Rumor mills: The social impact of rumor and legend*. Transaction Publishers, 2005.

[93] S. J. Frenda, R. M. Nichols, and E. F. Loftus, "Current Issues and Advances in Misinformation Research," *Current Directions in Psychological Science*, vol. 20, pp. 20–23, 2011.

[94] M. A. Hogg and D. L. Blaylock, *Extremism and the Psychology of Uncertainty*, vol. 8. John Wiley & Sons, 2011.

[95] B. Zhu, C. Chen, E. F. Loftus, C. Lin, Q. He, C. Chen, H. Li, R. K. Moyzis, J. Lessard, and Q. Dong, "Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities," *Personality and Individual Differences*, vol. 48, no. 8, pp. 889 – 894, 2010.

[96] A. A. Anderson, D. Brossard, D. A. Scheufele, M. A. Xenos, and P. Ladwig, "The nasty effect: Online incivility and risk perceptions of emerging technologies," *Journal of Computer-Mediated Communication*, vol. 19, no. 3, pp. 373–387, 2014.

[97] K. Coe, K. Kenski, and S. A. Rains, "Online and uncivil? Patterns and determinants of incivility in newspaper website comments," *Journal of Communication*, vol. 64, no. 4, pp. 658–679, 2014.

[98] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

[99] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N Project Report, Stanford*, pp. 1–12, 2009.

[100] T. T. Luong and D. Houston, "Public opinions of light rail service in Los Angeles, an analysis using Twitter data," *iConference 2015 Proceedings*, 2015.

[101] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining.," in *LREC*, vol. 10, pp. 1320–1326, 2010.

[102] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with Twitter: What 140 characters reveal about political sentiment.," *ICWSM*, vol. 10, pp. 178–185, 2010.

[103] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[104] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pp. 79–86, Association for Computational Linguistics, 2002.

[105] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts," *Journal of Artificial Intelligence Research*, pp. 723–762, 2014.

[106] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of Facebook 'friends': Social capital and college students' use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168, 2007.

[107] A. N. Joinson, "Looking at, looking up or keeping up with people?: motives and use of Facebook," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 1027–1036, ACM, 2008.

[108] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in Facebook," in *Proceedings of the 2nd ACM workshop on Online social networks*, pp. 37–42, ACM, 2009.

[109] Facebook, "Using the Graph API." Website, 8 2013. last checked: 09.05.2015.

[110] L. Gaudette and N. Japkowicz, "Evaluation methods for ordinal classification," in *Advances in Artificial Intelligence*, pp. 207–210, Springer, 2009.

[111] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.

[112] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[113] J. Cohen, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 70, no. 4, pp. 213–220, 1968.

[114] E. Frank and M. Hall, *A simple approach to ordinal classification*. Springer, 2001.

[115] B. Franks, A. Bangerter, and M. W. Bauer, "Conspiracy theories as quasi-religious mentality: an integrated account from cognitive science, social representations theory, and frame theory," *Frontiers in psychology*, vol. 4, 2013.

[116] C. Dewey, "What was fake on the Internet this week: Why this is the final column." The Washington Post, December 2015.

[117] C. R. Sunstein, "How Facebook Makes Us Dumber." Bloomberg View, January 2016.

[118] J. Brown, A. J. Broderick, and N. Lee, "Word of mouth communication within online communities: Conceptualizing the online social network," *Journal of interactive marketing*, vol. 21, no. 3, pp. 2–20, 2007.

[119] R. Kahn and D. Kellner, "New media and internet activism: from the 'battle of seattle' to blogging," *New media and society*, vol. 6, no. 1, pp. 87–95, 2004.

[120] R. Kumar, M. Mahdian, and M. McGlohon, "Dynamics of conversations," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, (New York, NY, USA), pp. 553–562, ACM, 2010.

[121] W. Quattrociocchi, "How does misinformation spread online?," *World Economic Forum*, 2015.

[122] W. Quattrociocchi, A. Scala, and C. R. Sunstein, "Echo chambers on Facebook," *Available at SSRN*, 2016.

[123] M. Takayasu, K. Sato, Y. Sano, K. Yamada, W. Miura, and H. Takayasu, "Rumor diffusion and convergence during the 3.11 earthquake: a Twitter case study," *PLoS one*, vol. 10, no. 4, p. e0121443, 2015.

[124] C. Betsch and K. Sachse, "Debunking vaccination myths: Strong risk negations can increase perceived vaccination risks.," *Health psychology*, vol. 32, no. 2, p. 146, 2013.

[125] A. A. AlMansour, L. Brankovic, and C. S. Iliopoulos, "A model for recalibrating credibility in different contexts and languages-a Twitter case study," *International Journal of Digital Information and Wireless Communications (IJDIWC)*, vol. 4, no. 1, pp. 53–62, 2014.

[126] G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, and A. Flammini, "Computational fact checking from knowledge networks," *arXiv preprint arXiv:1501.03471*, 2015.

[127] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, "Tweetcred: Real-time credibility assessment of content on Twitter," in *Social Informatics*, pp. 228–243, Springer, 2014.

[128] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, "Rumor has it: Identifying misinformation in microblogs," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1589–1599, Association for Computational Linguistics, 2011.

[129] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer, "Detecting and tracking political abuse in social media," in *ICWSM*, 2011.

[130] P. Resnick, S. Carton, S. Park, Y. Shen, and N. Zeffer, "Rumorlens: A system for analyzing the impact of rumors and corrections in social media," in *Proc. Computational Journalism Conference*, 2014.

[131] R. Peto and J. Peto, "Asymptotically efficient rank invariant test procedures," *J. Royal Statistical Society Ser. A*, no. 135, pp. 185–207, 1972.

[132] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society, Series B*, vol. 34, pp. 187–220, 1972.

[133] K. Krippendorff, *Content Analysis, An Introduction to Its Methodology*. Thousand Oaks, CA: Sage Publications, 3rd ed., 2012.

[134] I. Mozetič, M. Grčar, and J. Smailović, "Multilingual Twitter sentiment classification: The role of human annotators," *PLoS ONE*, vol. 11, pp. 1–26, 05 2016.

[135] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.

[136] A. Bessi, "Two samples test for discrete power-law distributions," 2015.

[137] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51(4), pp. 661–703, 2009.

[138] Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, vol. 57, 1989.

[139] A. M. McCright, R. E. Dunlap, and S. T. Marquart-Pyatt, "Political ideology and views about climate change in the european union," *Environmental Politics*, vol. 25, no. 2, pp. 338–358, 2016.

[140] R. E. Dunlap and A. M. McCright, "Challenging climate change," *Climate Change and Society: Sociological Perspectives*, p. 300, 2015.

[141] D. Elgesem, L. Steskal, and N. Diakopoulos, "Structure and content of the discourse on climate change in the blogosphere: The big picture," *Environmental Communication*, vol. 9, no. 2, pp. 169–188, 2015.

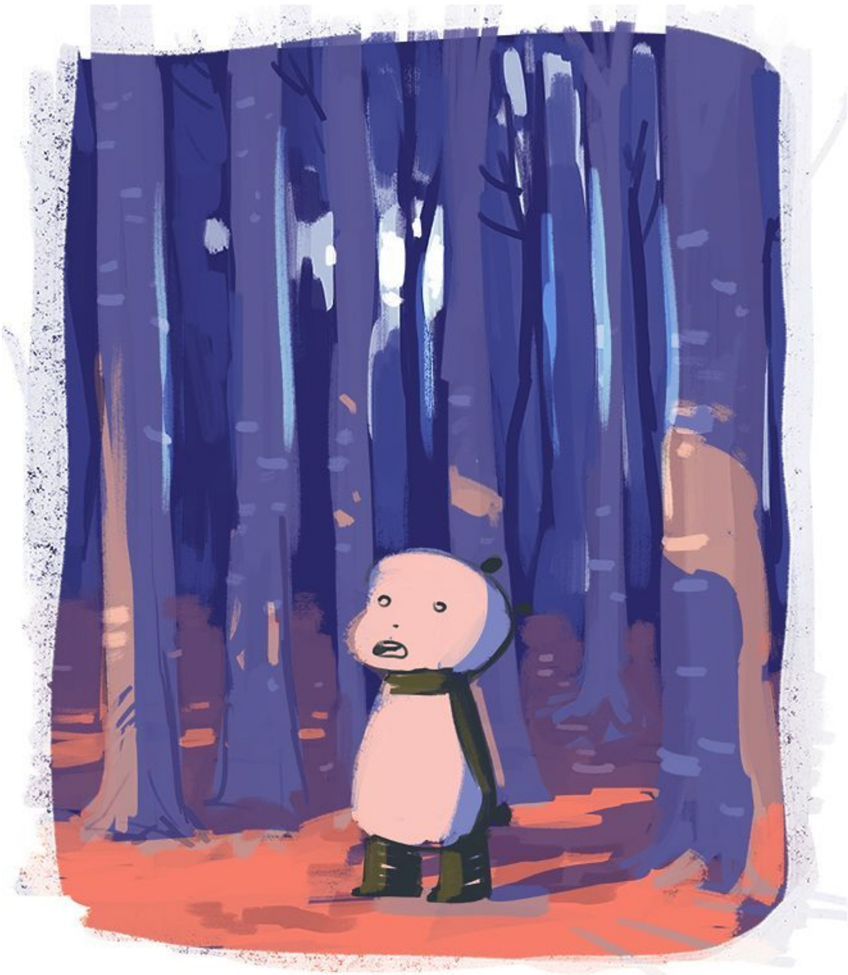
[142] N. J. Martin, J. L. Rice, and S. K. Lodhia, "Sustainable development planning: a case of public participation using online forums," *Sustainable Development*, vol. 22, no. 4, pp. 265–275, 2014.

[143] W. Pearce, K. Holmberg, I. Hellsten, and B. Nerlich, "Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 report," *PloS one*, vol. 9, no. 4, 2014.

[144] "Alchemy API — Powering the New AI Economy." An IBM Company, July 2016.

[145] A. Bessi, F. Zollo, M. Del Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, and W. Quattrociocchi, "Users polarization on Facebook and Youtube," *PLoS ONE*, vol. 11, pp. 1–24, 08 2016.

[146] A. Bessi, F. Petroni, M. Del Vicario, F. Zollo, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Homophily and polarization in the age of misinformation," *Eur. Phys. J. Special Topics (to appear)*, 2016.



-QUELLA LUCE ALLA FINE DELL'ANSIA-

FACEBOOK.COM/PANDAPIACE
©BEVILACQUA