UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

Scuola Politecnica e delle Scienze di Base
Department of Industrial Engineering (DII)

PhD in Industrial Engineering
XXXII Cycle

PhD Thesis

SOCIAL MEDIA EXPOSURE AND OPINION POLARIZATION: THE IMPACT OF GROUP IDEOLOGICAL DIVERSITY ON INTENTION FORMATION

SIMONETTA PRIMARIO

Tutor:
Prof. Giuseppe Zollo
Department of Industrial Engineering – University of Naples Federico II, Naples, IT

Co-Tutor:
Prof. Luca Iandoli,
College of Professional Studies – St. John’s University, New York, USA

March 2020

Contents

List of Figures

List of Tables

Embauché malgré moi dans l'usine à idées
j'ai refusé de pointer
Mobilisé de même dans l'armée des idées
j'ai déserté
Je n'ai jamais compris grand chose
Il n'y a jamais grand chose
ni petite chose
il y a autre chose
Autre chose
c'est ce que j'aime qui me plaît
et que je fais.

Jacques Prévert

To my Dad

who would surely be proud of me

ACKNOWLEDGEMENTS

Undertaking this Ph.D. has been a truly life-changing experience for me, and it would not have been possible to do without the support and guidance that I received from many people.

I am sincerely grateful to my tutors Prof. Giuseppe Zollo and Prof. Luca Iandoli, for involving me in such exciting research work, for their absolute trust in me, and for encouraged me to be professional and do the right thing, even when the road got tough. I could not have imagined having better mentors for my Ph.D. study.

I would like to thank Prof. Domenico Campisi, Prof. Gianluca Elia, and Prof. Simon Buckingham Shum for reviewing this thesis and providing interesting insights and comments.

I am also thankful to Prof. Carlo Lipizzi and Dario Borrelli for sharing their knowledge and technical know-how and for their constant support during these three years. Without their precious work, it would not be possible to conduct this research.

I thank you to all my co-authors, with a particular thought for Prof. Cristina Ponsiglione and Prof. Pierluigi Rippa, for the meetings and conversations that inspire me to think always outside the box.

I would also acknowledge all of my precious labmates for their feedback, cooperation, and friendship.

Finally, I would like to express my deepest gratitude to my family: my parents, my sister, and my boyfriend. This dissertation would not have been possible without their warm love, continued patience, and endless support.

ABSTRACT

At the beginning of its diffusion as a mass communication medium, the Internet had been enthusiastically saluted as a technology able to shift users from coordination models, based on hierarchical production and extrinsic market-driven incentives, to more effective, decentralized, collective, and intrinsically motivated peer to peer collaborations (Benkler, 2006; Castells, Fernández-Ardèvol, Qiu, & Sey, 2009; Levy, 1998; Malone & Klein, 2007; Tomlinson, 2007). Currently, several online platforms provide their users with opportunities and functionalities that support aggregation, membership, and interaction through connection, discussion, and content sharing. However, the literature on group thinking has consistently shown that interaction among members of a group can be conducive to dysfunctional dynamics that prejudice the development of individual and collective beliefs and choices. One of these dysfunctional dynamics is known whit the name of Group Polarization (GP) and refers to the tendency of groups to make decisions that are more extreme than the initial inclination of their members (Moscovici et al., 1972; Myers and Lamm, 1976). Over time, several studies have proved that GP can occur even when a group is not physically together, and that online groups are not immune to this form of group thinking, as well.

The research work presented in this thesis explores the role of SM on the development of this particular dysfunctional group dynamic and its implications on human behaviors. In particular, we focused our attention on the role of online diversity in the development of GP and peoples' intention to act in sustain a cause in which they believe. Overall, while our results seem to mitigate the worrying echo-chambers consequences, they demonstrate that it would be naïve to believe that greater diversity could attenuate the emergence of GP or surely improve decision-making.

INTRODUCTION

Research in behavioral economics has established that individuals tend to be subject to several biases that affect their decision-making. In this research area, the term *choice architecture* is used to indicate the design of different ways in which choices can be presented to people, and the impact of that presentation on their decision-making (Sunstein, 2001). A typical example of marketing choice architecture is the placement of retail products on shelves. The capacity of every element of this architecture to alter people's behavior predictably, without prohibiting the choice of alternatives or offering an incentive, is called a *nudge* (Thaler and Sunstein, 2009). For example, the conventional supermarkets' strategy "eye level is the buy level" indicates the "eye-level" as the nudge.

The design of effective nudges starts from the knowledge of the stakeholders and their interests and exploits cognitive and social mechanisms to favor one choice over another. The recent "datagate" about Cambridge Analytica has shown that the Internet could be a valuable ally to profile users' preferences and design personalized online experiences that silently favor some choices. Whether the Internet and Social media allow today to enhance the persuasive power of these nudges to make a user assume a behavior that otherwise would not have adopted, is the painful question of our time (Babaei et al., 2018; Pariser, 2011; Sunstein, 2002a; Thaler and Sunstein, 2009). In this thesis, we argue that the idea that the Internet can be such a threat deserves further empirical scrutiny given its diffusion and the regulatory implications that such a statement entails.

Purpose and research questions

This research work investigates the role of SM on the development of GP and peoples' intention to mobilize in sustaining a cause. Broadly defined, GP is the tendency by which an individual shifts his/her judgment towards more extreme positions under the influence of group interaction. The research questions that motivated this study can be summarized as follows:

RQ1: How do social media contribute to the emergence of group polarization?

RQ2: Are social media conducive to biased information consumption and suppression of diversity through the emergence of echo-chambers and information cocoons?

RQ3: Do social media contribute to making group interaction increasingly divisive and fragmented where each point of view is perceived as superior and non-negotiable?

Specifically, this work addresses at verifying:

- *Whether online exposure to polarized debate has an impact on people's beliefs and intentions formation*
- *What is the role of online diversity in the development of group polarization and peoples' intention to act.*

The chapters of this thesis are based on three research works that differently contributed to the purpose of the project. More specifically, in the first chapter, we defined the state of the art of literature about GP and online environment, providing hints about existing theories and methodologies, as well as research questions that are still open. After, we used this theoretical/methodological framework to set two empirical case studies. In the former, described in the second chapter, we studied GP on SM as an emerging dynamic led by the exposure of people to propagandistic and highly polarizing messages that came from different media (such as TV). While, in the second experiment, reported in the third chapter, we verified whether exposure to either diverse or homogeneous opinions in online discussions could alter people's stance on a controversial issue and their decision-making in the development of intention to act in sustain a cause in which they believe.

The positioning and contribution of the work

At the beginning of its diffusion as a mass communication medium, the Internet had been enthusiastically saluted as a technology able to shift users from coordination models, based on hierarchical production and extrinsic market-driven incentives, to more effective, decentralized, collective, and intrinsically motivated peer to peer collaborations

(Benkler, 2006; Castells et al., 2009; Levy, 1998; Malone and Klein, 2007; Tomlinson, 2007).

Currently, several online platforms provide their users with opportunities and functionalities that support aggregation, membership, and interaction through connection, discussion, and content sharing. However, the literature on group thinking has consistently shown that interaction among members of a group can be conducive to dysfunctional dynamics that prejudice the development of individual and collective beliefs and choices.

One of these dysfunctional dynamics is known whit the name of GP (GP) and refers to the tendency of groups to make decisions that are more extreme than the initial inclination of their members (Moscovici et al., 1972; Myers and Lamm, 1976). According to this definition, polarized groups make their members prone to higher risk if the individuals' initial tendency is to be risky, and towards greater caution, if individuals' initial tendency is to be cautious (Stoner, 1968).

Over time, several studies have proved that GP can occur even when a group is not physically together and that online groups are not immune to this form of group thinking, as well.

In particular, a group of Cyber-skeptics, led by Cass Sunstein, started to suspect that GP could be easily enhanced in online spaces, especially through the emergence of echo-chambers. The Echo-chamber has been proposed as a metaphor to identify online environments in which information, ideas, and beliefs are amplified or reinforced by the communication and interaction of like-minded people and biased information (Garrett, 2009; Grömping, 2014; Sunstein, 2002a, 2002b, 2001; Weinberger, 2004) . Similar to armored rooms in which knowledge circulates without encountering different points of view and contamination with outside sources, echo-chambers were accused of becoming the perfect ecosystem in which people with similar ideas are luring each other in a crescendo of anger and even violence (Sunstein, 2002a). In his works, Sunstein speculates that the emergence of echo-chambers leads groups to polarize and that the consequent radicalization would push their members to mobilize and act (Sunstein, 2001). In this sense, for the author, extremism could be capable of exploiting these environments to recruit new members, promote crime, spread ideas, and even coordinate attack (Sunstein, 2002a).

However, later researches provided evidence that partially confirm this view. A less stark perspective shows that while there is no shortage of online groups with extreme and homogenous ideology, Internet and SM can facilitate significant exposure to alternative points of view, for instance by favoring access to diverse information through the strength of weak ties (Bakshy et al., 2015; Barberá, 2014; Grabowicz et al., 2012). These findings were echoed by more recent studies showing that online users tend to feed themselves information from a varied media procured through serendipitous search (Bruns, 2019; Dubois and Blank, 2018; Flaxman et al., 2016).

Counterintuitively, some of these studies also showed that the presence of diversity is not necessarily a guarantee for more moderated and critical discussions. Indeed, the presence of antagonist points of view can be in itself a cause of polarization since the users' need to preserve their ideological identity in contrast with their adversaries (Bail et al., 2018; Bakshy et al., 2015; Lee et al., 2014; Levendusky, 2013). Described as a backfire effect GP could arise out of heated debates between opposing factions, each of which claims the own point of view as superior and non-negotiable (Bail et al., 2018; Del Vicario et al., 2016b; Mercier and Sperber, 2011).

Several are the contributions that this thesis offers to the heated debate briefly described above.

First of all, to the best of our knowledge, there is no coherent account of the literature exploring the relationship between GP and SM. Thus, our systematic literature review provides the first account, offering a collection of proposed theories, methodologies, and findings. Through this contribution, we favored the research progress moving from the current state of mostly isolated case studies to a more mature stage where researchers can anchor new findings in the context of an established body of knowledge.

Additionally, most of the works in the literature demonstrated to investigate GP as network configuration, reducing the phenomenon to a mere segregation of online groups. Thus, another contribution offered by this thesis concerns the study of polarization as a process that evolves and manifests itself through a particular dynamic. Therefore, we offer empirical evidences of how: i) polarization can emerge as a reaction of users to propagandistic and highly polarizing stimuli that may not come from the web but also other media such as TV, ii) measures of network, traffic, content, and sentiment might differently contribute to study the dynamic of GP in online environments.

Finally, considering that the presence or absence of a transfer effect from online to real action could magnify or downsize concerns about the consequences of SM exposure, we proposed a study to investigate how GP and peoples' intention to mobilize, emerge and evolve. And in particular, how SM influence people's decision-making enabling interaction among like-minded people and people with different points of view.

The evolution of the research

In order to describe the entire research project, this thesis is divided into four chapters. The first three chapters are based on some academic papers co-authored with other researchers, while the last one outlines general conclusions and future developments of this research project.

More in detail, in Chapter 1, we explored the academic literature through a systematic review process. This review provided an account of 92 works, offering a collection of theories, methodologies, and findings concerning GP and online environments. Overall, the results of this review showed that although the causes and mechanisms proposed to explain the online group dynamic do not far different from those that guide the phenomenon in face-to-face interaction, the online environment seems to offer more opportunities to design polarization-conducive environments. SM simultaneously offer open discussion environments where different points of view interact by confront and collide, and environments where interactions involve only like-minded people and beliefs circulate without encountering different points of view and contamination with outside views. However, both environments can offer valuable contributions to the emergence of GP and its consequences. Therefore, despite the numerous studies have addressed the impact of the online context on the phenomenon of GP, the understanding of this relationship is still in an early phase, where some claims are not empirically warranted and might have been exaggerated. At the same time, for others, there is no conclusive finding yet.

According to this review, we note that while there is a large number of studies that have shown that online communities tend to be quite segregated in terms of political orientation (Conover et al., 2011b; Garimella and Weber, 2017; Shore et al., 2018), less attention has been devoted to the analysis of the processes through which members of each faction talk to each other (online conversations). Moreover, most available research was about single medium studies, which makes hard to compare the consequences of exposure to specific information beyond the online medium being studied. Therefore, in Chapter 2, we intended to contribute to this debate by using a conversational analysis perspective applied to online political discussions and, in particular, to those where SM are used as a parallel channel where discuss about TV or live events (backchanneling). In particular, we analyzed 71,835 Twitter data related to the three U.S. presidential debates of 2016, and we measured GP through the use of a measure proposed by Morales et al., (2015). Our results showed that participants did get exposure to alternative points of view through conversational interaction with members of the opposite faction, but also that they reverted to interaction with like-minded peers after the debate was over. After with the help of data-mining techniques, we wondered which factors could better explain the evolution of this dynamic, and we discovered that GP is better explained when semantic metrics related to the content and structure of the conversation (that express who is talking about what) are combined with social metrics concerning the links between the participants (that express who talks with whom).

Finally, in Chapter 3, we proposed a controlled experimental design and the use of the well-known theory of planned behavior (Ajzen, 1991), to investigate whether exposure to either diverse or homogeneous opinions in online discussions could affect people's stance on an issue and their intention to act in sustain that. Specifically, we divided 356 subjects into 20 groups; ten involved only like-minded people, and ten involved people with different points of view. After a one-week online group discussion, we tested for differences across conditions and groups in terms of opinion shifts as well as intention to act and its precursors. Our statistics have shown that groups involving members with different points of view became more radical in terms of overall opinion, attitude, and subjective norms. Instead, although groups involved like-minded people reinforced both subjective norms and perceived behavioral control, they do not become more extreme in their position and attitude toward the behavior. Overall, these results offer empirical ground to support the concerns on the polarizing power of Internet-enabled debate and show that radicalization takes place in nuanced ways that, however, do not seem to make participants more likely to act.

Finally, the thesis concludes, in Chapter 4, with a general discussion on the limits, improvements, and further developments of this research project.

CHAPTER

ONE

ARE SOCIAL MEDIA DAMAGING THE QUALITY OF ONLINE DEBATE? A SYSTEMATIC REVIEW ON GROUP POLARIZATION IN ONLINE SETTINGS

Summary

Social media (SM) are often accused of worsening the quality of online debate. In this chapter, in particular, we focus on dysfunctional online group dynamics leading to opinion polarization and choice shift by which group participants may become more extreme in their initial position on an issue. Through a systematic literature review of the available research, we distilled a sample of 92 top journals and conference papers investigating polarization dynamics in SM and Internet-mediated communication, to summarize the main empirical findings and examine the available theoretical and methodological approaches. We use this knowledge base to assess six main accusations against SM in terms of their supposed tendency to amplify the causes and effects of GP. Based on this assessment, it appears that some of these claims are not empirically warranted and might have been exaggerated, while for others, there is no conclusive finding yet. On the base of this analysis, we propose a research agenda to address existing gaps and potential research directions.

Introduction

This chapter proposes a systematic review of the literature on GP in online discussions. Broadly speaking, GP refers to changes in individual attitude or behavior induced by peer pressure that makes the group members more extreme in their position on a controversial issue.

Studies on offline groups showed that members of polarized crowds could make their members prone to greater risk if the individuals' initial tendency is to be risky, and towards greater caution, if individuals' initial tendency is to be cautious (Aronson et al., 2002; Stoner, 1968). DiMaggio and colleagues (1996) define GP as both a state and a process. As a state, GP can be measured in terms of the extent to which opinions on an issue are opposed; as a process, GP refers to the increase in such divide over time (DiMaggio et al., 1996).

GP has received renewed attention following the increasing importance of SM in political communication, as highlighted by their supposedly critical role for diverse political movements and events such as the Arab Spring (Khondker, 2011; Wiest, 2011; Wolfsfeld et al., 2013), the Honk-Kong protests (Lee et al., 2017; Lee and Chan, 2016), the US elections in 2012 (Bor, 2014; Shin et al., 2017; Zhang et al., 2013) and 2016 (Enli, 2017; Groshek and Koc-Michalska, 2017), the rise of so-called populist parties in Europe (Ceron et al., 2014; Della Porta and Mosca, 2005; Gonzalez-Bailon et al., 2012; Treré et al., 2017; Vaccari et al., 2015), and the Cambridge Analytica/Facebook scandal (Assibong et al., 2020)

The academic community had warned against some of these trends since the beginning of the 2000s, as showed by several studies according to which online discussions, especially when taking place via SM, would favor extremization and antagonist debate that have a negative impact on the quality of political discourse and, ultimately, on democracy.

GP has been considered as a key product of online information cocoons and echo-chambers in which like-minded peers become more extreme in their positions through mutual reinforcement and encouragement (Sunstein, 2002a, 2001), and as one of the factors contributing to the diffusion of fake news and misinformation (Del Vicario et al., 2016a; Törnberg, 2018). These works stand in stark contrast with the initial optimism with which the Internet was enthusiastically saluted as a technology fostering democracy and collective intelligence in the pre-SM era (Levy, 1998; Malone and Klein, 2007). The online collaboration was expected to make peer to peer collaboration more efficient and effective than organized hierarchical production (Benkler, 2006; Castells et al., 2009; Tomlinson, 2007). Cyber-utopians advocated for a new Digital Renaissance in which more diversity, decentralization, and disintermediation would foster creativity, participation, and freedom (De Kerckhove, 1997).

Perhaps as a reaction to this view, contrarian voices started to raise concerns about the quality of the information that is created and shared online and, more broadly, on the impact of online media on democracy (Sunstein, 2001). Some cyber-skeptics argue that the impoverishment of the quality of online political discourse is inescapable since it stems from the understandable users' need to cope with information overflow. In order to help users to access relevant content, sophisticated personalization algorithms feed users with information that is consistent with their interests, preferences, and beliefs (Garrett, 2009; Stroud, 2010), which along with the human tendency towards homophily and self-confirmation biases, contribute to generate information bubbles (Pariser, 2012) leading to ideological closure, polarization, and radicalization (Garrett, 2009; Sunstein, 2002a)

This dispute seems far from being settled. More recently, cyber-skeptics conclusions have been rebutted by a new wave of studies offering evidence that information sharing overall produces more benefits than damage such as users' participation and exposure to diverse information (Barberá, 2014). Dubois & Blank, 2018 shows that online GP's negative consequences have been overstated, while other studies argue that there is no conclusive evidence that Internet makes online political debates more polarized than it is in society (Boxell et al., 2017a; Dilliplane, 2011; Prior, 2013). On the other hand, more and more anecdotal evidence is available showing that polarizing the debate to favor mobilization and persuasion of even a small number of undecided voters can be effective in tipping the balance in head-to-head competitions.

These contrasting findings, along with concerns about the emergence of GP and its effects, are some reasons that motivate this work. The second objective of this work is to review the theoretical models and the methodological approaches adopted to investigate the phenomenon since some of the empirical discrepancies can depend either on differences in the investigation methods and assumptions or on a lack of a more comprehensive understanding of online polarization. The methodological problem is even more pressing than ever since an increasing number of studies are based on the extraction of online analytics through data mining and other algorithms that are not always standardized or transparent enough to allow researchers to compare outcomes. Finally, to the best of our knowledge, there is no updated and systematic account of the existing literature and common findings of online GP.

The chapter is structured as follows. The methodology session describes the step by step analysis for the selection and the screening of the sources. In the results section, we first present bibliometrics extracted form a sample of the 92 top publications considered for this study. From this sample, we distill the main definitions, the theories, and the methodologies. Finally, we use this knowledge base to verify six accusations against SM impact on online GP and conclude our review with some research questions that are still open.

Methodology

Greenhalgh (1997) defined a systematic literature review (SLR) as an overview of primary studies that adopt clear and replicable methods to analyze significant contributions about a given topic comprehensively. Following well-established guidelines to SLR (Easterby-Smith et al., 2015) we started the process through a keyword search in two authoritative scientific databases (WoS and Scopus) and Google Scholar. The objective of the search was to identify a representative sample of most relevant academic works that have investigated GP in online discussions occurring through conversational platforms such as forums and SM. In particular, with conversational platforms, we refer to web-based technologies that facilitate multi-user interaction around the expression and creation of user-generated content, including the creation and sharing of information, ideas, or opinions via virtual communities and networks (Kietzmann et al., 2011). This definition includes platforms such as Facebook, Twitter, YouTube discussion, as well as blog comments sections and forums.

Our key search terms included online polarization, GP, opinion polarization, and echo-chamber in combination with the following terms: SM, social network, online platform, Web 2.0, blog, forum, and internet.

The articles obtained through this keywords string have been filtered through the following criteria:

- Disciplines: social sciences, computer science, multidisciplinary, psychology, neuroscience, business and management, decision sciences, behavioral science, and communications
- Language: only articles written in English
- Publication date: until December 2018
- Quality: articles published in peer-reviewed journals or peer-reviewed conference proceedings with a minimum of five citations for papers published before 2015 and at least one for more recent works

After eliminating duplicates. Titles and abstracts were screened to eliminate papers that did not focus on online GP. This step was performed independently by two of the three authors, and differences were resolved involving the third author.

A snowball search based on the references provided by works contained in this initial corpus allowed us to identify additional papers that had not been found through the initial database search.

The whole updated corpus was analyzed through the help of two software tools: NVivo and VOSviewer[1]. NVivo supports maintenance, annotation, classification, and encoding of key concepts, while VOSviewer provides tools for knowledge mapping and visualization based on bibliographic data and the content of the publications. Content Visualizations such as charts, word trees, and concept maps can assist by highlighting possible connections between works[2]. To achieve this objective, articles were organized in database records based on the following fields: Title, Author (s), Year, Source of publication, Number of citations, List of references, and PDF. The database can be interrogated through VOSviewer queries to obtain the distribution of works by years,

[1] https://www.qsrinternational.com/nvivo/nvivo-products/nvivo-12-plus
[2] https://www.vosviewer.com/download

publication source, and research area, and to create aggregated visuals. VOSviewer can be used to build networks in which nodes are objects of interest, such as publications, researchers, or terms, and links represent a relationship between them, such as co-authorship and co-citation.

NVivo supports the qualitative analysis of the content of the works in the corpus. NVivo can be used to search for specific phrases, words, or broad terms and quickly reveal the most frequently occurring words in these works. Moreover, the tool helps to produce cluster analysis of keywords to visualize patterns among concepts.

Results

The results of the systematic literature review are presented in the followed and organized into the following topics:

- Corpus definition and overview
- Online GP definitions
- Causal mechanisms, triggers, and effects
- Methodological issues

Corpus definition and overview

The initial keywords search on Scopus and WoS databases produced 906 results (520 Scopus and 386 WoS). The initial corpus was reduced to 596 papers (filtering by discipline) and then to 575 (filtering by language. After considering the only peer-reviewed journals and conferences, we eliminated an additional 12 articles. The exclusion of duplicates brought down the corpus to 439 works. Finally, we found only 204 papers published before 2015 that have received more than five citations and 94 works that have been quoted at least once after 2015. After reading the abstracts, 107 works were selected for full-text reading. After excluding 9 titles that were not available, a few more were excluded either because online interaction was not the focus (8 papers) or because they lacked a focus on GP (5 works). Finally, using Google Scholar, we identified an additional 7 papers through the analysis of the references of the remaining 85 papers. This process brought the final corpus up to 92 articles.

Early studies on polarization appeared in social psychology research already in the 1920s and ’30s (Bechterew and De Lange, 1924; Burtt, 1920; Thorndike, 1938). In an unpublished Master's thesis, MIT student James Stoner (1961) observed the so-called "risky shift," the tendency for individuals to propend towards riskier choices when they were in group decision-making situations than when they were alone (Stoner, 1961). The term polarization was introduced for the first time by Moscovici et al., in their analysis of the effects of group interaction on judgment and perception (Moscovici et al., 1972). Myers & Lamm (1976), reported that judgment polarization occurs both towards greater risks and greater cautiousness, depending on the initial orientation in the group majority (Myers and Lamm, 1976). These works, along with later review articles (Isenberg, 1986; Lamm, 1988), showed that polarization was determined by group informational influence on individual choices.

Online GP as a topic started to receive attention in the early 2000’s, through an influential paper by Cass Sunstein (2002), in which he defined GP as the tendency of “members of a deliberating group to predictably move toward a more extreme point in the direction indicated by the members’ pre-deliberation tendencies” (Sunstein, 2002b, p. 176). This work is among the first authoritative sources in connecting GP with online communication. Sunstein accuses Internet-enabled communication of favoring a social fragmentation into groups of like-minded-individuals and endorses Wallace’s hypothesis that “Internet like-setting is most likely to create a strong tendency toward GP when the members of the group feel some sense of group identity” (Wallace, 2001, p. 78)

As shown in fig. 1, works on online GP increases significantly thereafter and, markedly, after 2016, following the rise in research on the dark side of SM.

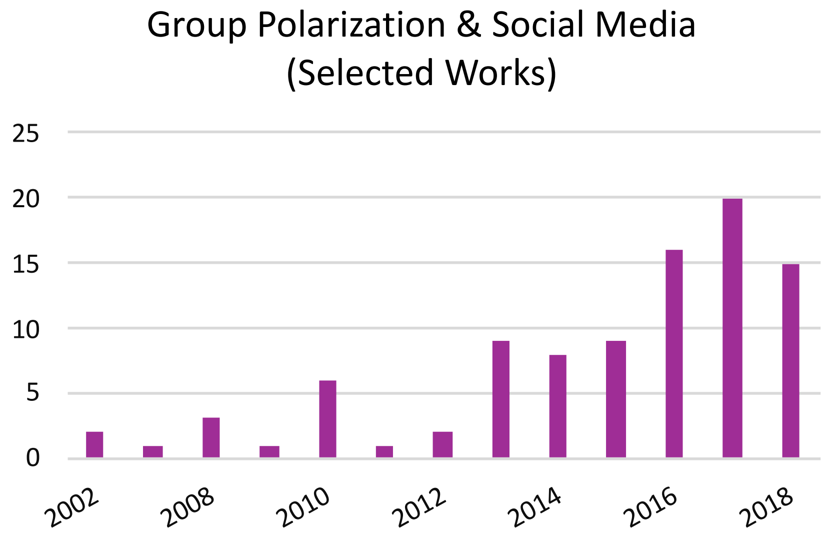


Figure 1: Number of works per year

As expected, we found that research in online GP is highly multidisciplinary (see fig. 2) and published in a variety of journals and conferences at the intersection between Computer and Social Studies (see tab. 1).

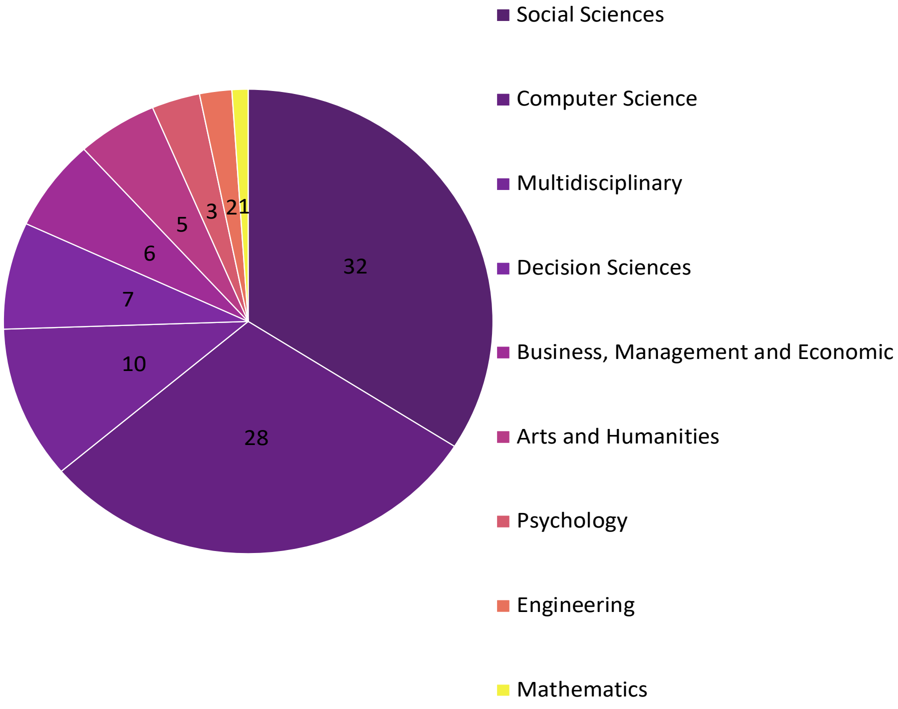


Figure 2: Research areas of selected works

Table 1: Source of publications

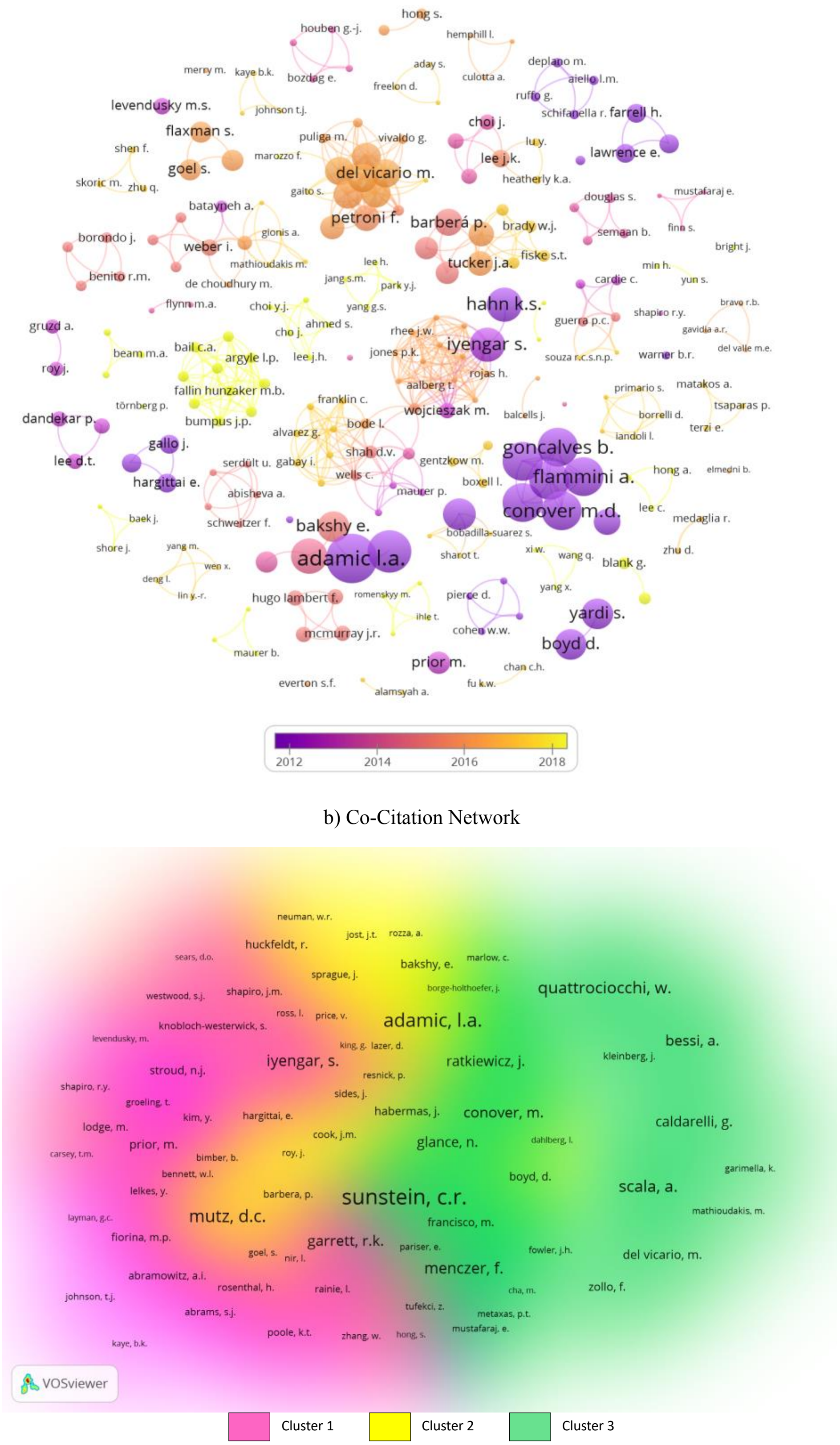
JOURNAL AND CONFERENCE PROCEEDINGS	#PAPERS
Association for Computing Machinery	7
Proceedings of the National Academy of Sciences	5
Journal of Communication	5
Institute of Electrical and Electronics Engineers	3
Journal of Computer-Mediated Communication	3
Public Choice	2
PLoS ONE	2
Scientific Reports	2
Communication Research	2
Policy and Internet	2
Computers in Human Behavior	2
Public Opinion Quarterly	2
Government Information Quarterly	2
New Media and Society	2

In order to complete the bibliometric analysis of the sample, we have built three different kinds of VOSviewer networks: co-authorship networks (fig. 3.a) and co-citation networks (fig. 3.b).

In fig. 3a, links indicate co-authorship. This network allows us to explore the collaborations among the 240 authors included in this review connected by 512 links distributed among 65 clusters. The size of each node indicates the number of citations, while colors indicate time of publication (more recent works in yellow and older ones in purple).

The graph shows the most cited research teams ordered by the time of publication of the most cited article. Table 2 reports these teams, most representative work, citations, and topics of interest.

Figure 3.b shows instead a co-citations graph, in which two or more nodes (authors) are connected to each other if there is at least one work that quotes them together. This graph was constructed by crossing the list of references of each work included in this revision. The entire network extracted through this analysis showed a total of 3958 authors co-cited at least once, but for visualization reasons, we reduced this network to 122 authors with at least ten co-citations among the selected works. However, this density visualization shows the presence of three clusters that include works frequently cited together. The green cluster, formed among works such as (Bessi et al., 2015; Conover et al., 2011a; Del Vicario et al., 2016a; Sunstein, 2002a), is the most numerous clusters and the one that collects the highest number of co-citations. Authors involved in this cluster analyzed GP as segregation of online communities in groups with high internal homogeneity. On the other side, in pink, we find a co-citations cluster among works such as (Garrett et al., 2014; Iyengar and Hahn, 2009; Stroud, 2010). According to these authors, GP is linked to online selective exposure and biased information content. Finally, in the middle, a yellow cluster (certainly less numerous) collects a higher number of citations through the contribution of works such as (Bakshy et al., 2015; Barberá, 2014; Mutz and Mondak, 2006) that investigate the degree of open discussions among groups with different views or exposed to heterogeneous information sources.


a) Co-Authorship Network

b) Co-Citation Network

Figure 3: Results of bibliographic network analyses

Table 2: Authors collaborations

Represented works	Citation	Main Topics
(Sunstein, 2008, 2002b, 2002a; Sunstein et al., 2017)	542*	Echo-chambers Radicalization Segregation Mobilization
(Adamic and Glance, 2005)	1109	Blog interactions Political blogs Presidential election Segregation
(Hargittai et al., 2008)	149	Cross-ideological discussions Blog interactions Political blogs Content of interactions
(Iyengar and Hahn, 2009)	644	Selective exposure Media consumption News topic Partisan media
(Yardi and Boyd, 2010)	223	Conversational heterogeneity Media Interactions In-group out-group affiliation Social corroboration
(Conover et al., 2011a)	876	Retweets and Mentions Community Structure Interaction analysis Network modularity
(Barberá, 2014; Barberá et al., 2015)	342*	Exposure to diversity Weak ties Discussion topic Ideological segregation
(Bakshy et al., 2015)	501	News topic Homophily Self-reported affiliation Content shared
(Bessi, 2016; Bessi et al., 2016b, 2016a, 2015; Del Vicario et al., 2017c, 2017a, 2016a, 2016b; Marozzo and Bessi, 2018)	450*	News consumption Echo-chamber Backfire effect Fake news
(Dubois and Blank, 2018)	50	Echo-chamber Media diversity High-choice environment Moderating media effect

*Sum of total citations

Using Nvivo’s 1-gram model functionality on titles and abstracts of the papers contained in the corpus, we created the tag cloud showed in the fig. 4 that reports the most common terms mentioned by the considered works. The cloud shows that online GP has been frequently investigated in the context of online politics and that Twitter and Facebook are the media that have received the most attention.

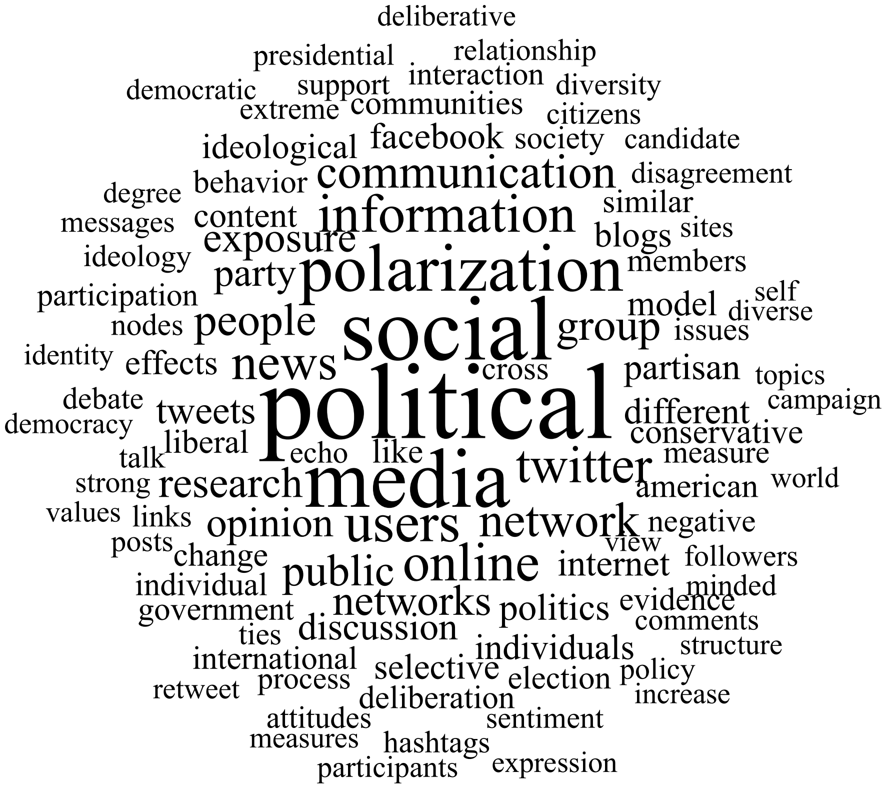


Figure 4: 1-gram graph, word cloud from title and abstract

Online GP definition: structure and process

While there is no general agreement on how authors define online GP (Tucker et al., 2018), existing literature appears to converge towards two fundamental aspects of online GP: segregation and radicalization, with some authors emphasizing one of these two traits over the other.

Studies focusing on segregation tend to consider online GP as the social process whereby a collective ends up being divided into antagonist sub-networks supporting irreconcilable, dichotomous positions, goals, and viewpoints (Balcells and Padró-Solanet, 2016; Guerra et al., 2013; Williams et al., 2015). Usually, this segregation reflects pre-existing political affiliation, as in Conservatives VS Liberals (Adamic and Glance, 2005; Bail et al., 2018; Primario et al., 2017; Stroud, 2010; Zhu et al., 2017) or about brand and religious preferences (Everton, 2016; Luo et al., 2013). Polarization also surfaces through value-based conversation on sensitive topics such as: gun-control (Brady et al., 2017; Garimella et al., 2017a, 2017b; Guerra et al., 2013; Merry, 2016), conspiracy theories (Bessi et al., 2016a, 2015; Del Vicario et al., 2016b), same-sex marriage (Barberá et al., 2015; Brady et al., 2017). In all of these cases, a population is polarized when it is divided into groups with opposite opinions (Garimella and Weber, 2017; Morales et al., 2015) with little or no communication and understanding of each other (Matakos et al., 2017). Lack of communication and connection can be visualized and, in part, quantified in terms of online social networks in which participants tend to communicate with or follow predominantly like-minded individuals.

Some authors have criticized this structural approach to the analysis of online GP by stating that social network segregation does not necessarily imply judgmental polarization (Everton, 2016; Levendusky, 2013; Morales et al., 2015). These authors define GP as the tendency for members of online groups to suppress internal diversity and align to the group's norm, typically assuming a more extreme position than they initially held under the group influence (Hong and Kim, 2016; Sunstein, 2002b, 2001; Warner, 2010). In these works polarization is considered as an escalation process in which individual opinions move toward either towards a more extreme point (Dandekar et al., 2013; Lee et al., 2014; Matakos et al., 2017; Romenskyy et al., 2018) or closer to the group shared norm (Sunstein, 2002a, 2002b, 2001). The emphasis on structural or process-based analysis of online GP has consequences on the way GP is measured and investigated. These operational aspects are presented in the following section in which we review methodological approaches.

Online group polarization: triggers, effects, and socio-psychological mechanisms

In the following, we classify the main findings of the research work on online GP included in our corpus using a framework based on three elements: mechanisms, triggers, and macro-effects. Mechanisms are causal explanations advanced in the literature of online GP based on social psychology theories. With triggers, we refer to conditions that facilitate the emergence of GP, and that appear to be particularly relevant in the online setting. Observable macro-effects refer to a set of macro phenomena that have been investigated as prominent products of online GP.

Mechanisms

Several theories have been proposed to explain the emergence of GP. Following a functional perspective, we define GP as a social dynamic through which the members of a group fulfill specific needs determining an individual judgment shift towards the group norm. By adopting insights from studies on organizational culture (Schein, 2010), we categorize these needs into two broad categories: internal cohesion and external adaptation. Theories such as Social comparison or Social identity belong to the first group while Persuasive Argumentation theory fits into the second group.

Social comparison (Festinger, 1954) can be offered to explain GP in terms of the need for individuals belonging to a group to conform. In this perspective, polarization is a judgment shift through which group members willingly align their opinions towards the perceived group norm (Stroud, 2010). Individuals may decide to change their preference and align to the group preference either to reduce cognitive dissonance when they believe the group is right and they are wrong or to increase social acceptance, whether the group is right or wrong. For Sunstein (2002a, 2002b), social corroboration, and positive reinforcement that individuals receive from the other members of their online group are the key processes that make online settings more prone to GP.

Social Identity Theory (Tajfel and Turner, 1979) explains polarization as the output of the process through which individuals reinforce their sense of membership to a social group. Self-categorization is the key process through which members of a group identify categories. After these groups are successfully recognized through the definition of more or less fuzzy boundaries, personal identity loses value in favor of a group identity, a simpler form of identification based on an "us VS them" rhetoric (Morin and Flynn, 2014; Suhay et al., 2018). The construction of group identity spreads both the feeling of not being alone and being different from others (Parsell, 2008; Suhay et al., 2018).

The construction of a social identity favors polarization since identity influences the way people assimilate information.

Turetsky and Riddle (2018) demonstrate that the use of a stereotypical linguistic frame in online networks modifies how people perceive and react to the news. While Shapiro (2013) shows that, when consuming online news, people tend to view an attitude-consistent news source as more believable, which, in turn, induces news' producers to bias their content in the direction of their audience. (Sunstein, 2002a) claims that self-categorization is enhanced in the online setting since the Internet increases both the variety of possible social categories (whatever your tastes are, you can always find groups of interest) and magnifies the visibility of marginal groups. Polarized individuals are likely to examine relevant evidence in a biased manner (Sunstein et al., 2017), and to accept confirmations at face value while subjecting dis-confirmations to critical evaluation (Dandekar et al., 2013).

Persuasive argumentation theory refers to GP as a mechanism through which a group exploits to prevail over other collectives perceived as antagonist or hostile (external adaption). Mercier and Sperber (2011) state that people are more likely to become radical in a position they support under the perceived urgency of defending this position from their opponents' attacks. SM would favor this dynamic in a few ways. First, these tools provide several affordances to construct and propagate narratives with strong emotional content through networks of allied individuals (Romenskyy et al., 2018). Second, they offer no native mechanism to prevent the diffusion of biased information (Sunstein, 2002a). Third, they favor exposure to alternative points of view, but this diversity backfires when participants have a preformed, biased position on a controversial issue fueling conflict through the provision of rhetorical online battlefields (Bail et al., 2018; Sunstein et al., 2017). Finally, biased but well-crafted narratives can shift the preferences of individuals that are neutral or undecided (Wojcieszak, 2010).

Persuasive argumentation theory can explain how slanted news outlets or provocative SM communication campaigns succeed in increasing polarization (Garrett et al., 2014; Wojcieszak, 2010). The same theory also helps to explain how motivated reasoning, i.e., people's tendency to produce arguments to support or rebut an idea depending on whether they agree with it or not (Prior, 2013): people do not interact with others to form an opinion since they already have one, but to win a rhetorical context against the proponents of the antagonist position.

In conclusion, available theories explain online polarization in different ways and posit that GP serves different needs. Approaches such as the ones based on SC and SI see GP as a confirmative process aimed at suppressing internal diversity to increase group cohesion and internal integration. GP manifests itself through a judgment shift towards the group norm, and polarization favors the recruitment of like-minded peers who escalate their beliefs through mutual reinforcement and support.

The argumentative approach sees polarization as an affirmative process through which members of a group fight rhetorical battles to prevail against their opponents. GP manifests itself through a judgment shift increasing the distance from the supporters of the rival position. Such a shift reduces internal diversity but magnifies the distance between the competing factions leading to heated debate and even verbal violence among participants of online conversations.

Based on the analysis of the works included in our corpus, we observed that theoretical accounts based on argumentative theories are more recent and minoritarian.

Triggers

Through the analysis of the papers contained in our corpus, we identified four main types of triggers of online GP:

- Selective Exposure and Homophily
- Cross-cutting exposure
- Polarized media and elite
- Events

While these triggers are conceptually distinct and favor GP in different ways, they can be at work and reinforce each other in online groups.

Selective Exposure. One way Internet users can cope with the information overload deriving from access to a variety of content and people is by using online groups and community as information filters (Del Vicario et al., 2016a; Parsell, 2008), through cognitive, social or algorithmic mechanisms (Bakshy et al., 2015; Pariser, 2012). Letting algorithms aside, depending on whether socio-cognitive filtering is about content or people, the trigger is, respectively, selective exposure or homophily. Selective exposure is about individual tendency to sift information and content based on pre-existing preferences or beliefs. Under the effect of this bias, and given the interactive nature of Internet media, online users will consume and diffuse content that reinforces their opinions and contribute to generate information cocoons through online interaction with other like-minded peers or information sources. In turn, support and consensus facilitate the formation of stronger and more extreme beliefs (Levendusky, 2013).

Homophily is people's tendency to gravitate towards participants that are perceived to be socially similar to themselves (Balcells and Padró-Solanet, 2016; Bessi et al., 2016a). SM make it easier for people to surround themselves with like-minded peers (Bessi et al., 2016a; Parsell, 2008; Williams et al., 2015). Some research findings link homophily to the fragmentation of the public into self-segregated communities (Lawrence et al., 2010; Medaglia and Zhu, 2017) that became a fertile breeding ground for GP (Wojcieszak, 2010). Mutual reinforcement, social support, and peer pressure originated in communities of similar minds will favor the emergence of stronger convincement.

Cross-cutting Exposure. Recent studies have argued against the idea that Internet media reduce diversity. These works recognize that echo-chambers and filter bubbles have been abundantly observed in experiments, but also claim that their importance in the real online world has been overstated (Balcells and Padró-Solanet, 2016; Bright, 2018; Garrett et al., 2014; Johnson et al., 2017; Lee et al., 2014; Shore et al., 2018; Weber et al., 2013). The most noticeable results obtained by these studies show that Internet users are intentionally serendipitous when searching for information (Semaan et al., 2014). Therefore, the level of diversity in the media diet for most users is quite high (Dubois and Blank, 2018). Finally, SM support as well the formation of weak ties that can help to increase the heterogeneity of the information and interaction (Barberá, 2014).

However, while it is true that the Internet offers its users the opportunity to encounter and engage with diverse individuals and information sources, this diversity can backfire in terms of the quality of online debates (Garrett et al., 2014; Yardi and Boyd, 2010). In such a context, factions supporting contrary opinions tend to collide, by trying to advance their motivations and emphasizing the difference between "us and them." In such an environment, people could fall in the so-called spiral of silence in which majoritarian or just louder opinions and sentiment become dominant and reduce to silence any minoritarian or dissenting alternative (Wells et al., 2017). Other studies have hypothesized that diversity can, in fact, increase opinion radicalization through heated debates occurring between opposite factions (Hargittai et al., 2008). Yardi and Boyd argue that, despite favoring exposure to a variety of point of views, SM tend to privilege haste and emotion (2010, p. 325).

Polarized media and elite. Some authors identify partisan media and biased elite as a fundamental trigger of online GP. The bias of these sources, along with their massive visibility and followership, favor polarization through the deliberate of divisive stories and arguments to increase the online source visibility (Iyengar and Hahn, 2009). Through the artful combination of biased narratives, strategic hyperlink connections, stereotypes, and clichés, polarizing media or sources can provide versions of facts that feed into the beliefs of the audience they are targeting an even monetize the online that generated by the public reaction (Messing and Westwood, 2014; Shapiro, 2013).

Divisive rhetoric aims at consolidating support from followers and at mobilizing undecided individuals through negative emotions and even outrage while attaching the opposing faction (Levendusky, 2013). Biased content spread from influential nodes to the online user community (Morales et al., 2015; Primario et al., 2017; Wells et al., 2017) via an intermediate layer of not necessarily biased other sources that end up magnifying the divide and grow emotional reactions (Lawrence et al., 2010; Prior, 2013; Stroud, 2010; Sunstein, 2002a).

Events. Finally, other authors hypothesize that the online GP reflects the divisiveness originated by polarizing events taking place offline, such as political elections, referendums, civil unrest, or other particular that solicit uproar and strong emotional reactions (e.g., violent crimes such as terrorist attacks). In these cases, SM can magnify discontent generated by emotionally intense events that can lead to outrage and overreaction. For instance, polarization associated with the gun control debate in the US is aggravated whenever a mass shooting occurs (Garimella et al., 2017b). In this case, the climate of conflict, distrust, and social malaise tend to strengthen beliefs and push people to take a side (Borge-Holthoefer et al., 2015; Gruzd and Roy, 2014; Romenskyy et al., 2018).

Effects of group polarization

While some scholars reported a few positive effects of online GP such as increase of trust in institutions and their representatives (Johnson et al., 2017) or the positive impact of brand polarization on customers enthusiasm and revenues (Luo et al., 2013), the analysis of the papers contained in our corpus reveals a dominance of adverse social effects, all associated to social fragmentation and sterile conflict:

- *Cyber-balkanization*
- *Radicalization and radical mobilization*
- *Diffusion of fake news and online misinformation*

Cyber-balkanization. The fragmentation of the online public sphere into sub-groups with specific interests (digital tribes) whose members predominantly interact with each other seems to be a recurring feature of the online social landscape (Chan and Fu, 2015; Lawrence et al., 2010; Wells et al., 2017). Cyber-balkanization refers to the special case in which members of online groups self-segregate into ideologically homogeneous communities that are shielded from dissent and have limited connection and interaction with other groups (Chan and Fu, 2017). Cyberbalkanization favors a state of isolation in which members of a group tend to increase ideological distance with outsiders (Bright, 2018; Heatherly et al., 2017), through biased choices of information sources (information cocoons) (Sunstein, 2002a) and of the participants with whom they interact (Adamic and Glance, 2005). Through a mechanism of selective avoidance, users shield themselves from undesirable or dissonant views by removing unwanted information and breaking social ties that transmit such information (Merry, 2016; Zhu et al., 2017). In turn, this tendency towards ideological closure farther nurtures online GP in a circular relationship of positive reinforcement (Stroud, 2010).

Radicalization and radical mobilization. Radicalization is defined as individuals' tendency to develop more extreme beliefs (Levendusky, 2013), feelings (Brady et al., 2017; Romenskyy et al., 2018), and attitudes (Weber et al., 2013). Polarized online groups tend to adopt violent forms of expression often linked to an extremist political, social, or religious ideology that challenges the established order (Garimella et al., 2016). Sunstein argues that online GP can lead to radicalization that, in turn, fuels violence and even terrorism (Sunstein, 2002b, 2002a). The charisma or persuasion of actors and messages within a group can contribute to accelerating further this dynamic (Everton, 2016) that can have detrimental societal impacts (Hong and Kim, 2016).

Some scholars also argue that opinion polarization can have an impact on offline choices and behaviors when it comes to mobilizing to support a given cause. Anecdotally, SM had a crucial role in the spreading and growing global protest from the web to the square (such as for the Arab Springs, the Metoo, the Friday for the future, and the Occupy Wall Street movements), thus enlarging the debate sphere on critical issues such as income inequality, climate change or gender discrimination. However, some authors offer evidence that polarized groups can equally be linked to the emergence of social tensions, turmoil, and violent clashes (Borge-Holthoefer et al., 2015; Lynch et al., 2017; Weber et al., 2013).

Fake news and misinformation. Several works included in this review see the dissemination of fake news and misinformation as another worrying social effect of online GP. Bessi and colleagues (2015) adopted online GP as a key metric to identify online groups in which false or misleading rumors were more likely to spread (Bessi et al., 2015). Polarized groups are more likely to welcome and spread information that is consistent with their and agenda their beliefs, regardless of whether this information is obtained from authoritative or more dubious sources. Introne et al. (2018), show that ideologically homogeneous communities weave facts into biased narratives leaning towards conspiracy stories. Interestingly, the same dynamic is observed in both pseudo-scientific and online scientific communities (Bessi et al., 2016a, 2016b, 2015; Del Vicario et al., 2016a). Törnberg describes the causal link between GP and misinformation with a captivating metaphor in which online GP has the same effect of a dry pile in a wildfire," [...] that provides the fuel for a small initial flame, and that can spread to larger sticks, branches, trees, to finally engulf the forest" (Törnberg, 2018, p. 2).

Methodological issues

Research methodologies adopted by the selected works are varied and differ by approach, measurements, and procedures.

Through the review of the selected works, we identified discussion topics, data sources (lab VS in the wild studies), time horizon, and polarization metrics. Figure 5 shows an overview of the results in terms of discussion content and data sources.

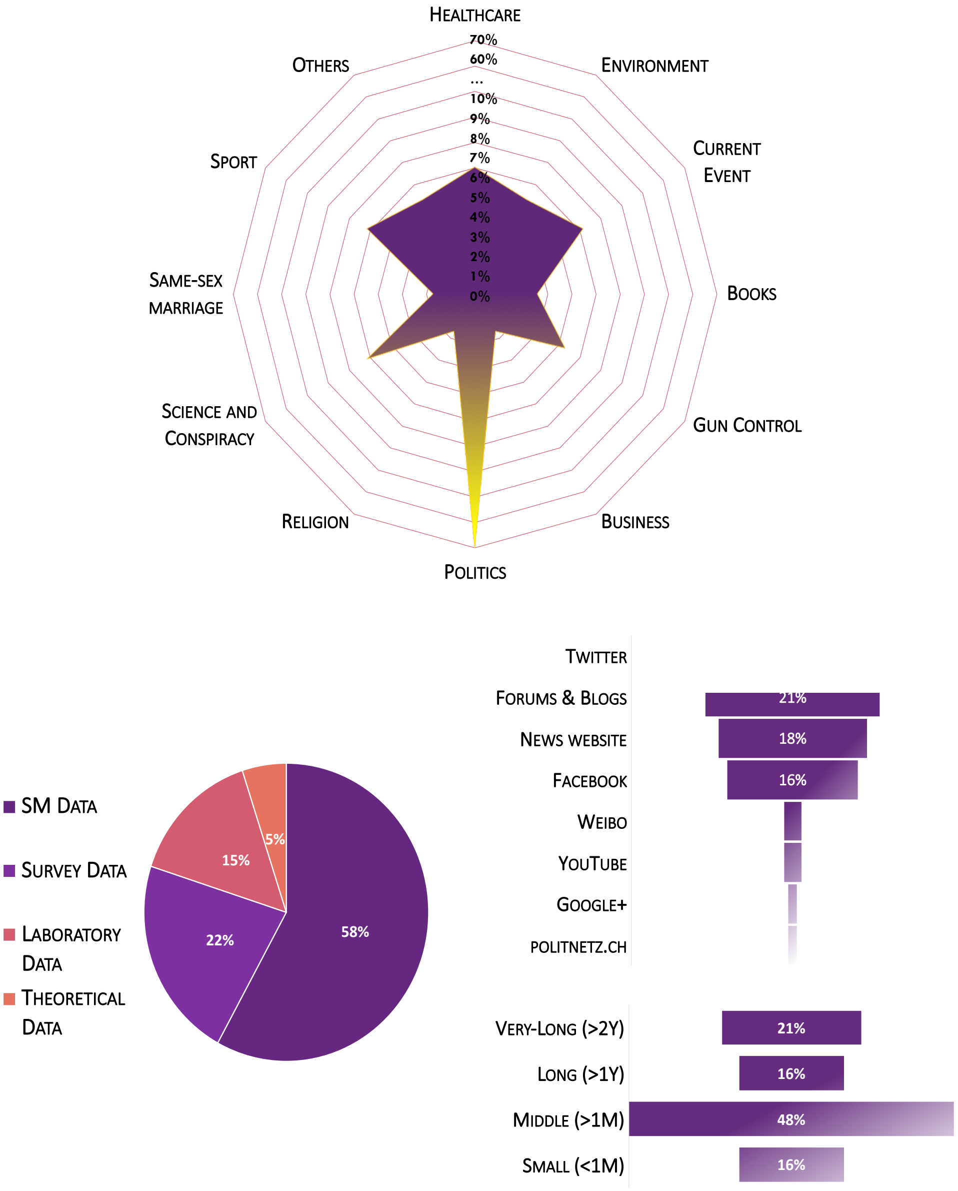


Figure 1: Overview of topic, data sources, SM, and temporal spaces of analyses

Topics of discussion. The spider web chart clearly shows that the domain that has received the most attention is online political discussions. Within the category Science and Conspiracy, we have included the six works conducted by the research group of Bessi, Del Vicario, Quattrociocchi, etc. that have studied the GP in relation to users involved in scientific/conspiracy discussions (Bessi et al., 2016a, 2016b, 2015; Del Vicario et al., 2017c, 2017a, 2016a). Polarization about health issues such as abortion (Cho et al., 2016; Garimella et al., 2017a, 2017b), diseases such as Ebola (Elmedni, 2016) or multiple disorders of personality (Parsell, 2008) have been classified into Healthcare. Instead, within Current Events we refer to works that have studied the polarization around events such as the death of Venezuelan president Hugo Chavez (Morales et al., 2015) the shooting of George Tiller (Yardi and Boyd, 2010) or that of Michael Brown in Ferguson (Park et al., 2018; Turetsky and Riddle, 2018). In general, most of the works investigate the phenomenon through the use of a single topic, while few (Barberá et al., 2015; Cho et al., 2016; Garimella et al., 2017a, 2017b; Iyengar and Hahn, 2009; Matakos et al., 2017) replicate their analysis over more topics. Most of the selected works use samples of data collected within a single country of origin.

Data sources. We classified data sources into 4 main categories: theoretical (metadata from literature reviews, conceptual works, and simulation data), Laboratory (data acquired in controlled experimental designs), Survey (data acquired through questionnaires), and SM (data collected from online platforms). While theoretical data sources (based on speculation and rational hypotheses) shown limits about their applicability and empirical coverage, survey data sources are often accused of being based on self-reporting that suffers from measurement errors. SM mining in the wild emerged as a response to self-report bias problem, providing a relatively complete and objective digital trace of individual opinions, attitudes, and activities. In turn, while SM mining can be derived to track online behavior, there is no guarantee that these findings can be generalized to offline behavior. Sometimes surveys and interviews were built specifically for the studies (Bail et al., 2018; Lee et al., 2014; Wojcieszak, 2010), others come from national surveys and authors extract the necessary information for the study of the factors under control (Boxell et al., 2017a; Parsons, 2010).

Most research uses Twitter as a source of SM data (39%), followed by forums and blogs (21%), online news websites (18%) and Facebook (15%). Data collection methods also differ in terms of observation duration, ranging from small (less than a month), medium (between one month and one year), or long (more than 2 years) periods. Finally, the data collection process is often consequential to the choice of source. While some works adopted manual transcription of messages/links (Cho et al., 2016; Turetsky and

Riddle, 2018), it is definitely more common to proceed through extensive data collection through code libraries and APIs.

Metrics. An overview with a description of the most used or cited metrics is reported in table 3 below. Several approaches and metrics have been proposed to assess online GP. Many works use indirect measures based on user perceptions or verbal escalations collected through feeling thermometers and Likert scales.

The alternative and increasingly more common approach is to measure GP directly from SM data, exploring the level of segregation between groups or their expressions.

Some works use social network analysis metrics in which nodes are users (Bozdag et al., 2014), elites (Garcia et al., 2015), or sources such as blogs or group pages (Adamic and Glance, 2005; Costa e Silva, 2014), and links map relations of friendship (Del Vicario et al., 2017a; Garcia et al., 2015; Medaglia and Zhu, 2017, 2016), or interactions like retweets and mentions (Adamic and Glance, 2005; Bravo et al., 2016; Conover et al., 2011a), co-retweets (Finn et al., 2014), or a combination of the above (Garcia et al., 2015; Williams et al., 2015). Network-based metrics aim at capturing network fragmentation and segregation as proxies for GP. Thus, authors exploit well-known network measures (i.e., modularity, Conover et al., 2011) or the degree of openness/interaction across two sub-communities (i.e., Community Boundaries, Guerra, et al., 2013 or the Network Fragmentation, Morales et al., 2015).

Network-based metrics have been criticized for several reasons. First, topological measures do not provide insight at the content level. Second, the presence of either fragmentation or interaction among segregated groups is not necessarily the effect of more polarized debate (Guerra et al., 2013; Primario et al., 2017; Yardi and Boyd, 2010).

Content-based metrics based on a semi or fully automated analyses have then been proposed to address some of the above limitations. Content metrics have been used to identify ideologically separate communities (Gruzd and Roy, 2014; Hemphill et al., 2016), the existence of antagonist narratives (Bode et al., 2015; Marozzo and Bessi, 2018; Turetsky and Riddle, 2018), the emergence of more extreme beliefs (Garimella et al., 2016; Romenskyy et al., 2018; Weber et al., 2013), or the relationship between online GP and online sentiment (Alamsyah and Adityawarman, 2017; Borge-Holthoefer et al., 2015; Finn et al., 2014; Gruzd and Roy, 2014; Merry, 2016; Primario et al., 2017).

Table 3: Most used GP measurements

MEASURE	DESCRIPTION	FORMULA	REFERENCE	INVESTIGATED LEVEL
By Feeling Thermometer	The feeling thermometer determines and allow to compare respondents' feelings about a given question, issue, or sentence. Based on this tool, authors enable respondents to express their beliefs by applying a numeric rating of their feelings to an imaginary scale. Respondents express their feelings in terms of degrees, with their attitudes corresponding to temperatures.	Polarization is measured by the difference between the feeling thermometer index of parties (i.e., Democrat and Republican), which ranges from 0 to 100. Then the first thermometer scores (Republican) were subtracted from the second thermometer scores (Democrat). The calculated absolute value represents the degree of polarization.	(Beam et al., 2018; Boxell et al., 2017a; Cho et al., 2016; Min and Yun, 2018; Suhay et al., 2018)	Perception Verbal escalation
By Likert scale	The scaling method measures either a positive or negative response to a statement. A Likert item is simply a statement that the respondent is asked to evaluate by giving it a quantitative value on any kind of subjective or objective dimension, with the level of agreement/disagreement being the dimension most commonly used.	Subjects' position on an issue was collected through a set of questions such as: "Do you think X is right or wrong"; "How do you feel thinking about X." Answers are measured on an n-items Likert scale with a central neutral answer ("I don't know"), surrounded by biased items on each side ("extremely favorable/extremely agree," "favorable/agree," "unfavorable/disagree," and "extremely unfavorable/ extremely disagree"). Then, polarization is measured considering the extreme answers or the distance from a neutral position.	(Dubois and Blank, 2018; Heatherly et al., 2017; Lee et al., 2018, 2014; Medaglia and Zhu, 2017, 2016; Wang et al., 2018; Warner, 2010; Wojcieszak, 2010; Yang et al., 2016; Zhu et al., 2017)	Perception Verbal escalation

Modularity	A measure of the structure of networks quantifies the quality of a partition in terms of a modular structure.	Given an unweighted network, modularity Q is defined as: where N and m are the amounts of nodes and links in the network, kiout and kiin are the out-degree and the in-degree of node i, Aij is the adjacency matrix of the network, and δ(i, j) is a function that takes the value 1 if nodes i and j are in the same group, and 0 otherwise. For the case of a weighted network, the amount of links is replaced with the sum of weights of all links, and the adjacency matrix has entries corresponding to the weight of each link.	(Alamsyah and Adityawarman, 2017; Borge-Holthoefer et al., 2015; Conover et al., 2011a; Del Vicario et al., 2017a; Garcia et al., 2015; Lynch et al., 2017; Shore et al., 2018; Turetsky and Riddle, 2018; Williams et al., 2015)	Network segregation Verbal escalation
Community Boundaries	A social structures that highlight the presence (or absence) of antagonism. The basic idea is that in a network there could be nodes that effectively interact with the (potentially) opposing group. These nodes are part of a community boundary. A polarized network is characterized by a small number of nodes preferring connection across their community boundary.	They calculate the polarization P of the network as: where: B = boundary; di(v) = the number of edges node v has in Eint, with v ∊ B; db(v) = the number of edges node v has in EB, with v ∊ B; Eint = the set of edges that connect boundary nodes to internal nodes; EB = the set of edges that connect members from Gi to members from Gj P lies in the range (-1/2,+1/2). A P value below 0 indicates not only lack of polarization, but also that nodes in the boundary are more likely to connect to the other side. While, a P value greater than 0 indicates that, on average, nodes on the boundary tend to connect to internal nodes rather than to nodes from the other group, indicating that antagonism is likely to be present.	(Bright, 2018; Guerra et al., 2017, 2013)	Network segregation

Bimodality Coefficient	Statistical measure that indicates the presence of modal values in a distribution of opinions. An uniform or exponential distribution shows a bimodality coefficient equal to 5/9. Values greater than 5/9 may indicate a bimodal or multimodal distribution or even heavily skewed unimodal distributions. The maximum value (1) is reached only by a Bernoulli distribution with only two distinct values.	The formula for a finite sample is with µ3 referring to the skewness of the distribution and µ4 referring to its excess kurtosis, with both moments being corrected for sample bias using the sample size n.	(Bessi et al., 2016a, 2016b; Matakos et al., 2017; Romenskyy et al., 2018)	Network segregation
Network Fragmentation	A measure of the system's overall polarity inspired by the electric dipole. In the simplest case of two point charges of opposite signs (–q and +q), like the electric dipole moment, this measure is proportional to the distance among the charges. Therefore, the network is perfectly polarized when it is divided in two groups of the same size and opposite opinions.	Given a network of users, the network fragmentation index is measured as: Where is the difference in size between both populations and the pole distance, d, measuring as the normalized distance between the two gravity centers.	(Morales et al., 2015; Primario et al., 2017)	Network segregation

Network Polarity	The ratio of sum of edges contained within each group over the total number of edges present in the network.	Let Ei be the number of edges connecting nodes within group i and Eij be the number of edges connecting nodes between groups i and j. Then, in a network containing k groups, its polarity degree is:	(Finn et al., 2014)	Network segregation Verbal escalation
Random Walk Controversy	Consider two random walks, one ending in partition X and one ending in partition Y, this index quantify the difference of the probabilities of two events: (i) both random walks started from the partition they ended in and (ii) both random walks started in a partition other than the one they ended in.	The measure is quantified as Where ,A,B∈{X,Y} is the conditional probability	(Garimella et al., 2017a, 2016; Garimella and Weber, 2017; Weber et al., 2013)	Network segregation
Network Disagreement	The degree of disagreement between users' expressed opinions, measured in different moment. This measure considers polarization as a process according to which people's opinions changing their entity since externally influenced.	Given a graph G = (V, E, w) and a vector of opinion x ∈ [0, 1]n of individuals in V, the network disagreement index $\eta(G, x)$ is defined as $\eta:=\in$ Consider an opinion formation process over a network G = (V, E, w) that transforms a set of initial opinions x ∈ [0, 1]n into a set of opinion x'∈ [0, 1]n. The process is polarizing if $\eta(G, x') > \eta(G, x)$, and vice versa.	(Dandekar et al., 2013; Flaxman et al., 2016)	Network segregation Verbal escalation

DW-NOMINATE scores	A scaling political measure representing legislators on a spatial map. In this sense, a spatial map is much like a road map— the closeness of two legislators on the map shows how similar their voting records are. Using this measure of distance, DW-NOMINATE is able to recover the "dimensions" that inform congressional voting behavior. The primary dimension through common pole "liberal" vs. "conservative" (also referred to as "left" vs. "right"). A second dimension picks up differences within the major political parties over slavery, currency, nativism, civil rights, and lifestyle issues during history.	The linear DW-Nominate scores are constructed as follows: The two variables, and , represent conservative and liberal ideological indexes for politician i, respectively, while DWNOM is the first dimension of the DW-Nominate score. To calculate the value, took the average value of the DW-Nominate score for all the politicians [mean(DW Nominate Score) in the two prior equations and then subtracted this average value from each individual ideology score. Finally, took the absolute value; thus, 0 indicates a moderate position, whereas a positive number indicates more politically extreme or distinct positions (either conservative or liberal).	(Hong and Kim, 2016; Lawrence et al., 2010)	Network segregation
Sentiment Polarity	An approximation of the emotional attitude of users towards one piece of information that they shared by considering the sentiment of the text.	Labeled the sentiment of each comment as: negative/con (− 1), neutral (0), or positive/pro (+ 1), sentiment polarization (ρσ(i)) as it follows: where Ni, ki, hi are respectively the number of all, negative, and neutral comments left by user i, while li = Ni − ki − hi is the number of the positive ones. Note that ∈ [− 1, 1] and that it is equal to 0 if and only if li = ki or hi = Ni, it is equal to 1 if and only if ki = Ni, and it is equal to − 1 if and only if li = Ni.	(Del Vicario et al., 2016b; Lee et al., 2014; Williams et al., 2015)	Network segregation Verbal escalation

Discussion

In order to summarize the main findings and identify possible gaps and open research questions, we identified 6 claims that are often advanced by both the scientific literature and by mainstream media regarding SM's role in favoring the emergence of polarized online debate. The first 3 claims regard the triggers of online GP, while the other 3 points are related to empirical evidence about the effects of GP (table 4).

Table 4: Claims and empirical evidences

CLAIM	ASSESSMENT
1. SM disproportionately favor selective exposure and like-minded interactions that enable GP	Conflicting evidence
2. SM offer argumentative spaces where supporters of different ideologies collide and polarize	Wide consensus
3. SM is a powerful tool that is deliberately used by partisan media and elite to fuel GP	Wide consensus
4. Online debate enabled by SM is overall highly fragmented and polarized	Conflicting evidence
5. GP eases the diffusion of fake news and misinformation through SM	Some evidence
6. Online GP favors offline mobilization	Limited evidence

1. Social media disproportionately favor selective exposure and like-minded interactions that enable group polarization

Sunstein provided several theoretical arguments to support this statement (Sunstein, 2008, 2002b, 2002a). Subsequent works provided empirical evidence supporting this accusation (Iyengar and Hahn, 2009; Lawrence et al., 2010; Medaglia and Yang, 2017; Medaglia and Zhu, 2016; Stroud, 2010; Warner, 2010; Wojcieszak, 2010). However, other works show that homophily and selective exposure are to ascribe to individual choices more than to the characteristics of SM (Bakshy et al., 2015; Lee et al., 2014; Messing and Westwood, 2014); that people tend to consume selectively political information online just as they do offline (Shapiro, 2013); and that the effectiveness of exposure to diversity and its absence seems to be strongly dependent on the strength of people's pre-existing beliefs (Bright, 2018; Dandekar et al., 2013; Johnson et al., 2017).

Equally numerous are the works that do not find empirical support for Sunstein's accusation. According to Dubois et al., 2018 the hypothesis underlying this conclusion is overestimated since it is based on studies on single media and narrow definitions (Dubois and Blank, 2018). Flaxman and colleagues show that although there is an increase in the ideological distance between individuals, this is not associated with the existence of filter bubbles or echo chambers since the results show cross-ideological interactions (Flaxman et al., 2016). Other studies demonstrate that users are more likely to be exposed to disagreement (Beam et al., 2018; Dubois and Blank, 2018) by choice (Semaan et al., 2014) or thanks to the existence of weak ties that increase the probability of encountering different opinions (Barberá, 2014).

2. Social media offer argumentative spaces where supporters of different ideologies collide and polarize

Based on our analysis, there is a strong consensus that exposure to diversity ends up increasing online GP. SM provide opportunities to directly engage one's opponents, to instigate a virtual fight through online confrontation (Merry, 2016). Bail and colleagues observed that after one month, Republicans who followed a liberal Twitter bot became substantially more conservative, while Democrats exhibited slight increases in liberal attitudes after following a conservative Twitter bot (Bail et al., 2018). Cross-minded interactions allow Twitter users to reinforce in-group and out-group affiliation (Morin and Flynn, 2014; Yardi and Boyd, 2010). Williams and colleagues confirm that SM discussions on climate change occurring cross ideologies carry out a stronger negative sentiment (Williams et al., 2015). Bode et al. (2015) found empirical evidence that self-affiliated users and organized online communities strategically exploit Twitter conversation to align themselves to identities, contexts, and media of their choice (Bode et al., 2015). This alignment can also take place through co-retweeting actions (Finn et al., 2014). Described as a backfire effect, criticism encountered online allows people to develop a greater trust towards parties to which they feel they belong to, and more repudiation opponents (Suhay et al., 2018). Finally, as Zhu and his colleagues point out, selective exposure and homophily can be the consequence of exposure to diversity. The construction of isolated environments such as those described in the first accusation can be the result of a prior condition of cross exposure that the users themselves refuse or avoid (Zhu et al., 2017).

A minority of scholars, however, consider SM-enabled debate as potentially beneficial. Balcells and Padró-Solanet (2016), first show that cross-minded conversations tend to be significantly longer than like-minded ones, which can lead to more genuine and articulated deliberation. SM, such as Facebook and Twitter, can offer valuable democratic spaces that can favor de-polarization (Beam et al., 2018) or that anyway do not lead to increase polarization beyond pre-existing levels (Merry, 2016). Parsons (2010) observes that disagreements can actually depolarize emotions by decreasing negative attitudes toward a candidate of the opposite party. Many of the comments analyzed in Beam, and colleagues' study indeed showed that users engaging in cross-minded conversations recognized in whom had a different opinion not as just an enemy but a valid interlocutor with whom to discuss (Beam et al., 2018; Gruzd and Roy, 2014). Overall, not all SM encourage the same level of exposure to diversity and political polarization, and the characteristics of the platforms can lead to differentiated effects (Min and Yun, 2018). However, personal characteristics of users seem to be the decisive factor over the technological features of the communication platform being used (Heatherly et al., 2017).

3. Social media is a powerful tool that is deliberately used by partisan media and Elite to fuel group polarization

Levendusky (2013) work shows that motivated reasoning and persuasive arguments of partisan media can polarize the electorate. The strategic use of hashtags, keywords, and hyperlinks supports the accusation of an induced GP. The spread of stereotypes and distorted narratives was strongly polarizing in the case analyzed by (Turetsky and Riddle, 2018). This study demonstrated that GP occurred when media sources were more likely to link selectively linking to news coverage that shared the same emotional valence and stereotype-relevant aspects of the events. Moreover, not only news sources but also politicians, community leaders, influencers (that we referred to as Elite) express moral emotions—of either positive or negative valence—in an effort to increase message exposure and to polarize perceived norms (Brady et al., 2017). Morales et al. (2015) propose a model in which opinion spreads through a process of influence from Elite nodes to intermediate nodes of opinion leaders, up to the great audience of listeners, Strong partisan news media were also found around the Italian referendum, and in this circumstance, their support for one or the other political party remained unchanged over time (Marozzo and Bessi, 2018). In a comparative study, Yang et al. (2017) found that for all of 10 investigated counties, people who acquired their news online tended to perceive a more polarized polity, and if they had extreme issue positions, they also perceive more polarization among the parties.

Some authors do not agree with this conclusion. Prior (2013) concludes his meticulous review stating that "there is no firm evidence that partisan media are making ordinary Americans more partisan". Bessi suggests a model that emphasizes the importance of specific users' personality traits in the choice of endorsing narratives proposed by polarizing media (Bessi, 2016).

Yang et al. (2017) noted that the influential effect of partisan media and Elite affects the perception of polarization rather than the attitude polarization. Boxell et al. (2017) reported findings showing that political polarization was strong among population segments, which are not likely to use SM.

4. Online debate enabled by SM is overall highly fragmented and polarized

Most of the works included in this review have linked the phenomenon of GP to cyberbalkanization, a tendency towards isolation and radicalization of online communities (Chan and Fu, 2017). Adamic and Glange (2005) offer one first empirical evidence of cyberbalkanization among American blogs. This work showed that online blogs strongly segregated into ideological clusters (Republicans and Democrats) with few links between clusters of different ideologies. Two different situations emerge around the cyberbalkanization. Ideological minorities are unlikely to emerge in such a highly cyberbalkanized environment (Bozdag et al., 2014). Bessi et al. (2015) also found that Facebook users tend to be very polarized with respect to science or conspiracy subjects, by forming distinct groups characterized by strong homophily and similar information consumption patterns. Cyberbalkanization also includes a process of opposition between clusters through the development and exploitation of a uniform, simple and biased language (Garimella et al., 2017b; Hemphill et al., 2016; Weber et al., 2013).

When was observed over the long term, cyberbalkanization exhibit strong peaks close to specific events or circumstances (such as elections, catastrophes, or political debates), followed by more or less flat valleys (Garcia et al., 2015; Garimella and Weber, 2017; Hanna et al., 2013; Primario et al., 2017; Yang et al., 2017). Furthermore, analyzing different types of networks, it was observed that the generation of a more cyberbalkanized environment was more evident in retweet and follower networks, rather than in mentions networks (Bravo et al., 2016; Conover et al., 2011a; Williams et al., 2015). Signs of openness between the different communities have also been linked to new Twitter features, such as Quote Retweets (Garimella et al., 2016). Although these results have led researchers to consider the retweet as a declaration of endorsement, the results of Guerra et al. (2017) show that this assumption could lead to incorrect conclusions. In particular, the authors show that there may be cases in which retweets can be used to cite the creator of original content out of the original context of the message, for derision or criticism.

The accusation of cyberbalkanization does not find support also in the analyses conducted by Costa e Silva about the Portuguese political blogosphere. In this work, different opinions were confronted through civil and constructive discourse, without polarization (Costa e Silva, 2014). Furthermore, signs of interaction among users with opposite or undefined orientation, refute the cyberbalkanization emergence in the work of (Alamsyah and Adityawarman, 2017).

5. Group polarization eases the diffusion of fake news and misinformation through social media

Online GP facilitates the diffusion of information of questionable quality (Bessi et al., 2016a, 2015; Del Vicario et al., 2016a). The basic idea is that easy access to unlimited but biased information and contents create overconfident people that reject expert advice because they presume to be knowledgeable enough (Parsell, 2008). Even worse, experts are considered by many as an elite group with vested interests, sometimes acting to serve the power or even to accomplish a hidden agenda (as in conspiracy communities) (Bessi et al., 2015; Marozzo and Bessi, 2018). Online groups easier to accept information that confirms their beliefs, even if they contain deliberately false statements (Bessi et al., 2016a; Del Vicario et al., 2017b, 2016a). In this context, polarization is defined as the user's preference between two antagonist narratives (e.g., scientific VS and conspirational). Törnberg (2018) shows how the simple condition of homophily may not be able to explain fake news virality patterns, adding polarization as a driver element.

The dissemination of misinformation can be considered the effect of a worrying mechanism in which false news, that if presented in a coherent way to some ideologies, not only end up being more easily accepted by the community but also develop a greater intention to the subsequent re-sharing of the news. Given the dependence between the truthfulness attributed to the news and the volume of arguments and approvals, this reinforcement could be fatal (Wang et al., 2018).

6. Online Group polarization favors offline mobilizations

In a famous article titled "Why they hate us," written in the aftermath of the 9/11 attack, Cass Sunstein hypothesized a direct relationship between the emergence of online echo chambers and political radicalization (Sunstein, 2002a). According to Sunstein, SM becomes the perfect ecosystem in which people with similar ideas are luring each other in a crescendo of anger and even violence. Violent extremism is effective in exploiting these environments to recruit new members, promote crime, spread ideas, and even mobilize and coordinate attacks. This capitalization of online polarization leads the extremism to mobilization and action. This last point aims to understand how dangerous targeted agitation can be when modern information warfare techniques back it. In such context emotions of hearths make polarized world views increasingly irreconcilable (Romenskyy et al., 2018). Although some authors do not always investigate the causal relationship between polarization-induced radicalization and the numerous mobilizations and actions of violence. Weber et al. (2013) proved that the level of polarization, in its verbal manifestation through tweets, can be used to forecast the unexpected outbreak of violence occurred in Egypt in late November 2012. In a Ukrainian case study, Romenskyy found that the opinion split may facilitate the separatist trends on its own (Romenskyy et al., 2018). Actions such as military coups reached popular support thanks to the political conflict and the spread of online fear and hatred, which undermined the democratic transition in Egypt (Lynch et al., 2017). Other empirical evidence demonstrated that it is far more likely for GP to manifest itself violently if groups believe that the use of violence is divinely sanctioned (Everton, 2016). Finally, Elmedni's (2016) study noted that SM can play an effective role in increasing the existing polarization and thus distancing public decisions from the rational model. This can take the form of action (as in the case of the political response to Ebola) or inaction (as in the situation of climate change and tax reforms). Both action and inaction are forms of public choices that do not necessarily serve particular public interests (Elmedni, 2016)

Conclusion

In this systematic review of the literature, we have collected 92 works that aimed to shed light on the impact that SM have on the GP dynamics. Understanding this relationship is an important issue for scholars, digital entrepreneurs, and regulators alike, given the potential effects that online GP can have in terms of compromising democracy, promoting circulation and the proliferation of false news, and mobilizations. Despite the fact that numerous studies have addressed this phenomenon, the understanding of this relationship is still in an early phase.

As our descriptive statistics of bibliographic data shown, the importance of the phenomenon and its relevance has caused an increasing interest and the adoption of a variety of theoretical and methodological approaches to study this group dynamic. We analyzed these approaches, highlighting what definitions, theories, and methodologies have been adopted in order to investigate. Finally, 6 claims regarding the supposed negative role of SM in facilitating GP have been assessed.

Despite some contrasting findings, overall, our analysis was not able to find a direct and clear impact of SM technological platforms as a cause for GP.

GP instead appears to be the consequence of: i) individual choices in the way content is consumed on SM; ii) the way SM are strategically used and possibly manipulated by polarizing actors; iii) the prevalence of revenues models that are driven by traffic, regardless of the quality of the content; and iv) typical social dynamics that are at work both online and offline.

On the other hand, it is safe to admit that SM platforms do not provide any countermeasure or disincentive to polarizing group dynamics and that, because of their scale and popularity, they magnify to a much larger scale this phenomenon so that its perception ends up being stronger than what is warranted by empirical evidence.

Finally, online debate, especially when occurring on SM, lacks the necessary level of transparency in terms of the users' awareness of being exposed to biased information. While affiliation to a group and selection of information is still the user's choice, there are many ways through which this exposure can be subtly and obscurely manipulated through algorithmic solutions such as bots or filtering mechanisms.

We conclude this review with a series of research questions that remain open, and can act as an agenda for future research.

Based on our discussion, it could be interesting for future works to investigate online GP in a more analytical fashion, for instance, by developing models describing *how the polarization process unfolds and progresses through different stages*. Equally interesting would be to know more about the impact of GP on individual *intention to pursue certain choices and behavior* in order to assess if polarization just translates into heated debates or has more concrete effects. Anecdotal evidence, such as the one offered in the Cambridge Analytica scandal, shows that polarizing content can tip voters' choices; more empirical studies are needed to confirm this finding. It would also be interesting to investigate *under which conditions such influence can take place*, for instance, whether polarization influences action only after a certain critical threshold is passed.

Additional other research is also needed to identify effective countermeasures. How can we identify *incentives or design platforms that spot and penalize polarized debates* without disrupting social interaction? *Is some content more conducive to biased thinking?*

At the theoretical level, our review shows that theories inspired by persuasive argumentation provide a better explanation of why GP can thrive online. We argue that a more in-depth analysis of argumentation strategies and dynamics in online conversation could not only improve our understanding of the phenomenon but also potentially provide insights for the design of better platforms and incentives to support higher-quality debate without limiting participation and freedom of speech. We think there is a concerning trend of growing consensus around proposals aimed at limiting and controlling user-generated content, often driven by a limited understanding of the phenomenon and a rational assessment of its consequences. We hope this work contribute to these aims and could be useful to other scholars in the pursuit of a better but always open and free Internet.

CHAPTER TWO

A CONVERSATIONAL PERSPECTIVE ON THE ANALYSIS OF IDEOLOGICAL SEGREGATION IN SOCIAL MEDIA-ENABLED POLITICAL DEBATE

Summary

Existing literature has shown that the emergence of communities of like-minded members due to opinion polarization in political debate is a recurring feature in the internet political landscape. However, less attention has been devoted to the conversational processes that occur within these communities when SM are used as a parallel channel where discuss about TV or live events (backchanneling). In this study, we analyze Twitter data collected during the Presidential television debates in the last 2016 US general elections. Our results show that participants did get exposure to alternative points of view through conversational interaction with members of the opposite faction, but also that they reverted to interaction with like-minded peers after the debate was over. The model shows that this event-driven fragmentation is better explained when semantic metrics related to the content and conversational structure (who is talking about what) are combined with social metrics based on the links between the participants (who talks with whom). As a complement to previous work ascribing network fragmentation to social factors such as homophily and social identity, we conclude that conversational interaction also reflects this partisanship despite the exposure to diverse information and audience.

Introduction

In this chapter, we intend to contribute to the debate introduced above by using a conversational analysis perspective applied to online political discussions enabled by SM among members of opposed political leaning in two main ways:

> *- While there is a large number of studies that have shown that online communities tend to be quite segregated in terms of political orientation, less attention has been devoted to the analysis of the very processes through which members of these factions talk to each other: online conversations;*
>
> *- Most available research is about single medium studies, which makes it hard to compare the consequences of exposure to specific information beyond the online medium being studied. In this chapter, we study Twitter-enabled interaction coupled with the use of another medium, the television in our case (backchanneling). Thus, here we analyze online dynamics regarding what happens in the offline world.*

Regarding the first point, we argue that the analysis of the conversational micro-interactions among participants can help to improve our understanding of how fragmentation into like-minded groups works and is generated. In particular, we are interested in quantifying how much of this interaction occurs across segregated communities and whether potential exposure to diverse information strengthens or dilutes partisanship.

In order to answer these questions, we tracked the evolution of Twitter conversations taking place around television political debates occurring at the same time. By observing the Twitter reply network, we found that conversational interaction did occur across antagonist political communities, but that this interaction was temporary and induced by the TV event, while interaction taking place between like-minded participants seemed to dominate before and after the event itself. Our results also show that this exposure to different points of view through the interaction with members of the other group produced adversarial conversations that were not likely to moderate participants' initial political leaning. Finally, using data-mining techniques, we identified a model showing that this adversarial dynamic was driven more by the content and structure of the conversation as determined by the polarizing TV debate than by social network metrics and sentiment. Based on these results, we speculate that the characteristics of the medium may have a role in favoring conversational dynamics that, in the long term, contribute to inflate an "us VS them" rhetoric and to strengthen selective exposure and social identity that facilitate partisanship and ideologically polarized reasoning. After presenting the empirical results, we discuss the research implications and limitations of this study and outline ideas for additional research.

Background and research questions

Several works have provided evidence that the Internet political landscape is fragmented into online communities populated by individuals with similar political leaning. One cause has been identified in selective exposure to information that is consistent with pre-existing beliefs. Selective exposure can occur either because people actively prefer to consume information that is supportive of their preferences (confirmation bias) or because collaborative filtering algorithms and other media affordances can be engineered to sift out information that is not aligned with the user's tastes (Pariser, 2012).

Other scholars have identified homophily and social identity as other causes for online fragmentation) (Settle, 2019; Wojcieszak, 2010). Homophily refers to the tendency to bond with similar others based on various factors, including gender, age, race, status, religion, political ideology. Social identity theory predicts that the membership can strongly shape individual identity to a specific group. The availability of like-minded weak ties created through SM helps to strengthen the impression that many people can support an idea or perspective we endorse, and that therefore those perspectives are valid. Homophily and social identity can help to generate isolated "echo chambers" that favor fragmentation of the public into communities of self-segregated collectives that end up to be exposed to highly biased information (Lawrence et al., 2010; Medaglia and Yang, 2017; Van Alstyne and Brynjolfsson, 2005) and whose members tend to become more extreme toward the preferred position (Sobkowicz and Sobkowicz, 2012).

Another cause for opinion-driven network fragmentation is self-categorization (Turner and Reynolds, 2011). According to this theory, individuals identify with a particular group and conform to the dominant group position. While self-categorization does happen offline, SM reduces the costs of both reaching and affiliating to groups based on specific interests while helping in the creation of a much larger audience (Twenge et al., 2016).

Other works show how the aggregation of individuals into groups of like-minded peers can favor the emergence of other adverse effects such as the diffusion of false information (Bessi et al., 2015; Du and Gregory, 2016; Törnberg, 2018) or the adoption of violent and aggressive tones in online discussions (Alamsyah and Adityawarman, 2017; Habibi et al., 2014; Lai et al., 2015; Williams et al., 2015).

Our review reveals some interesting, open, and widely debated questions. First, Can online network ideological fragmentation be a critical factor in explaining online behavior, especially in those cases in which online interaction becomes dysfunctional? Second, do SM and the Internet, in general, favor the emergence of ideological fragmentation and online partisanship because of the ways these media are designed and online interaction monetized? Third, Can online interaction balance the trend towards homogeneity and closure via the exposure and access to diverse and multiple sources? Fourth, is there any significant and durable impact of online fragmentation on individual behavior and choice in the online and even in the offline world? For instance, are SM a cause of political radicalization and mobilization towards extremism, as stated by (Sunstein, 2001)?

In this chapter, we intend to contribute to the study of the emergence of online partisan networks in the contest of online political discourse by positioning our work in the debate related to the second and third questions. Additionally, we intend to add to the paucity of studies focusing on the role of content (Bail et al., 2018; Barberá, 2014; Hargittai et al., 2008), in particular when created through online conversations, as opposed to works on online networks that have predominantly focused on the analysis of the social determinants of fragmentation. More specifically, we intend to address the following issues:

- Does the exposure to contrasting information and opinions via SM favor or hinder the emergence of these networks in online political debate?

- Which processes determine online fragmentation and partisanship in SM? More specifically, how conversational interaction takes place within and across politically homogeneous online groups and how it affects their emergence?

A conversational analysis approach to the investigation of online political partisanship

In this chapter, we adopt a conversational metaphor for the analysis of backchanneling twitter streams and show that this metaphor can provide useful conceptual and methodological tools to improve our understanding of content creation and its effects in SM streams.

According to the Common Ground theory, a conversation is a form of collective action requiring participants to coordinate on content and the process (Clark and Brennan, 1991). Content coordination is achieved when participants build a shared understanding of what they are talking about through the accumulation of mutual knowledge defined as common ground. Process Coordination is accomplished through conversational turns regulating participants' access to the conversation (Clark and Schaefer, 1989). A critical output of any conversation is the updating of the common ground, through the assessment and acknowledgment of new contributions.

We assume that in backchanneling applications, Twitter streams are loosely coupled conversations: the strong focus on the specific event being followed helps participants to accomplish content coordination (as opposed to free broadcasting that is not driven by a specific event). At the same time, the use of Twitter functions such as "mentions," "retweets," and "replies" supports some process coordination through the creation of a basic reply structure.

The common ground theory provides a set of concepts that can be used to model user-driven online content generation in SM.

First, Twitter conversations may lead to common ground accumulation; however, since the coordination on the process is weak, content negotiation and validation are problematic, especially when users have alternative and conflicting stances on the issue under discussion. Coherent accumulation is a possible indicator that can help to distinguish less polarized conversations from a more fragmented one.

Second, the concept of adjacency pairs (Clark and Schaefer, 1989) can also be exploited. Any conversation can be represented as a flow of adjacency pairs uttered within the turn-taking structure.

In a Twitter conversation, adjacency pairs are made possible through a few dialogical functionalities such as retweet, reply, and mention that allow participants to "reply" to each other. In the proposed methodology, adjacency pairs are used in both logic and dialogic roles. The dialogic role helps to reconstruct the evolution of a conversation through the analysis of the various reply moves in Twitter (i.e., who "said" what to whom). In its logic role, adjacency is reinterpreted as semantic proximity to favor common ground accumulation: two contributions are linked because they semantically overlap; two users are linked because they are talking about the same object. As we show in the next section, both the logic and dialogical roles of adjacency pairs can help to map Twitter streams into dynamics network of keywords, a formal representation from which a variety of structural and semantic metrics can be computed.

In the context of Twitter, we consider the available reply affordances (e.g., retweet, mention, and reply-to functions) as means users can exploit to support the construction of adjacency pairs. The limited space available, however, does not provide space for re-elaboration and content negotiation so that adjacency pairs can be used primarily for two purposes: endorse or attack someone else position.

For instance, participants can retweet a message they agree or disagree with, possibly followed by some supporting or derogatory short comment.

The dominance of adjacency pairs that reiterate either the support of or the attacking to a given point of view or position is predicted by the Persuasive Arguments Theory (Mercier and Sperber, 2011). According to this theory, individuals, during debates, tend to reflect less on the content and more on how to defend their attitude. Consequently, attachments to certain beliefs will push individuals to be more biased and confrontational towards people having alternative beliefs and to aggregate around those who support their positions.

Based on the common ground and the persuasive argument theories, we hypothesize that since content coordination is problematic given the limited affordances of the medium, conversational interaction among members of opposing groups will be event-driven and predominantly adversarial. In other words, we predict that individuals will increase their engagement with peers leaning towards opposite beliefs during the event as opposed to before and after the event itself either to attack the opponent's positions or to support their preferred position. We also hypothesize that the fragmentation of the conversational network will be predominantly driven by the semantics, and the structure, of the exchanged content. More specifically, we expect that conversations happening among like-minded peers will be more cohesive and diverse than those taking place among participants belonging to opposite factions, since participants with a similar background and political orientation will more easily build common ground.

Methodology

In this section, we adopt a methodology developed by (Lipizzi et al., 2016) based on common ground theory and persuasive argument theory to extract three types of variables from Twitter streams:

1. the conversational network based on the Twitter reply functions (Who talks to whom);

2. the content network represented as a concept maps built from adjacency pairs (What users are talking about);

3. a metric derived to identify the degree of fragmentation of the conversational network based on conflicting opinions (Do participants talk to each other across groups with opposite stances?).

Finally, using data mining techniques, we build a model to explain network fragmentation based on content and social network metrics extracted from the conversational and the content networks. The methodology is articulated in the following steps:

1. Data-collection and pre-processing.

2. Concept-map extraction from the conversation.

3. Computation of semantic and social metrics.

4. Analysis of network fragmentation based on Twitter reply structure

This method is implemented through a Python procedure and combines 11 scripts, for a total of about 1200 lines of code (see Appendix A).

Data collection and pre-processing

Tweets were download via the Twitter search API through multiple downloads and stored in a non-structured database (MongoDB). Consequently, tweets are randomly sampled by the API from the Twitter pipeline. Search parameters need to be defined (e.g., keywords) along with other filters such as language or country. Each tweet was then imported into a structured database as a triple ⟨u, W, t⟩ where $u \in U$ is a user sending the tweet, W is a vector of words w_i composing the text of the tweet, t is the time the tweet has been posted.

In the pre-processing step, conversational contents are summarized into a list of keywords. This output is performed through an automated text mining Python script and the use of Natural Language Toolkit and CLIPS/pattern through the following steps, via tokenization, elimination of stop words, lemmatization, and identification of n-grams (e.g., words that co-occur very frequently in the context and need to be treated as a single entity such as "green energy" or "President Obama"). The list of keywords is then manually inspected and further cleaned by duplications typically due to misspelling, as well as from other irrelevant and spam elements.

The conversation is then split into time slices of variable duration. A segmentation criterion needs to be established based on the objective of the analysis (see Clauset and Eagle, (2012) for guidelines).

Concept map extraction

In this step, we perform for each slice a dynamic reconstruction of "what people are talking about" based on the identification of the logic adjacency pairs contained in the analyzed Twitter stream using a method based on bipartite networks. Bipartite networks (also called "2- mode" networks) associate elements of one set to elements of a different set.

In our case, the two disjoint sets are the set of users, denoted with U, and the set of keywords, indicated with V.

A Twitter stream can be represented through a bipartite graph G = (U, V, A) such that if u_i is a Twitter user and v_j is a keyword, there is an edge $a_{ij} = (u_i, v_j) \in$ A if and only if u_i sent a tweet containing the keyword v_j.

In order to obtain the network of indirect relationships between nodes in a same set, we need to create a 1-mode network composed by nodes from one of the two sets U or V, linked together if they have at least one neighbor in common in the other set in G. Because our goal is to analyze conversational content, our 1-mode network will be based on the set V (words). The V-projection of the G graph is:

$$G_V = G \times G_T = (V, A_V), \text{ where } A_V = \{(p, q), \exists\, x \in U: (p, x) \in A \text{ and } (x, q) \in A\}$$

In other words, G_V represents a network of words that are connected because they are "shared" by users directly or through some degree of separation. Those structures are "socially generated" and they include logic adjacency pairs obtained through the concatenation of keywords linked together by different users.

Computation of semantic and social metrics

In this step we compute a set of metrics to analyze the concept maps in terms of semantic and topological properties. These metrics include:

1. Lexical diversity (number of unique keywords in the text divided by the total number of keywords in the text)
2. Sentiment polarity based on SentiStrength (Thelwall, Haustein, Larivière, and Sugimoto, 2013).
3. Topological metrics computed over both the keywords and the reply networks including total number of edges and nodes in the graph, betweenness centrality, and network density. More cohesive keyword networks will be denser and more structured (e.g. cluster topics connected by hub keywords bridging two or more different topics)

All the above metrics are collected for each time partition.

Network fragmentation metric

Following Morales et al. (2015), a network fragmentation metric is computed starting from the Twitter reply network and elites' opinions (the authors refer to this metric as network polarization). Nodes are classified into two groups: S (elite nodes) and L (listeners). Listeners' opinions (X_l) are initially set to a neutral value (zero), and each listener l updates his/her opinion at time t based on the mean opinion valence of his/her neighbors j:

$$X_l(t) = \frac{\sum_j A_{lj}\, X_j(t-1)}{k_l^{out}}$$

where:

A_{lj} = element of the adjacency matrix equal to 1 if there's a link between l and j; otherwise A_{lj} =0

K_{lout} =outdegree of node l

-1 ≤ X_l ≤ 1, with -1 and 1 representing two opposite preferences (e.g. two alternative candidates).

Elites' opinions (X_S) are instead supposed to be biased towards either choice, and are not affected by influence from the network. The computation of the preference is performed through several iterations until there is a convergence in the listeners' opinion. If the network is polarized around two conflicting opinions (e.g., coded with +1 and -1), the density distribution of the preferences p(X) is bimodal, and the higher the polarization, the further apart the peaks will be (fig.6). Opinion fragmentation is calculated as the momentum of an electric dipole, as follows. First, the size A_+ and A_- of the polarized population are computed:

$$A^+ = \int_0^{+1} p(X)dX = P(X>0) \quad ; \quad A^- = \int_{-1}^{0} p(X)dX = P(X<0)$$

$$\Delta A = |A^+ - A^-| = |P(X>0) - P(X<0)|$$

In order to measure the distance between the two populations the center of gravity of each distribution is determined, along with their distance *d*:

$$gc^{+} = \frac{\int_{0}^{+1} p(X)XdX}{\int_{0}^{+1} p(X)dX}; \qquad gc^{-} = \frac{\int_{-1}^{0} p(X)XdX}{\int_{-1}^{0} p(X)dX};$$

$$d = \frac{|gc^{+} - gc^{-}|}{|X_{max} - X_{min}|} = \frac{|gc^{+} - gc^{-}|}{2}$$

Finally, network fragmentation (μ), is computed as follows:

$$\mu = (1 - \Delta A)$$

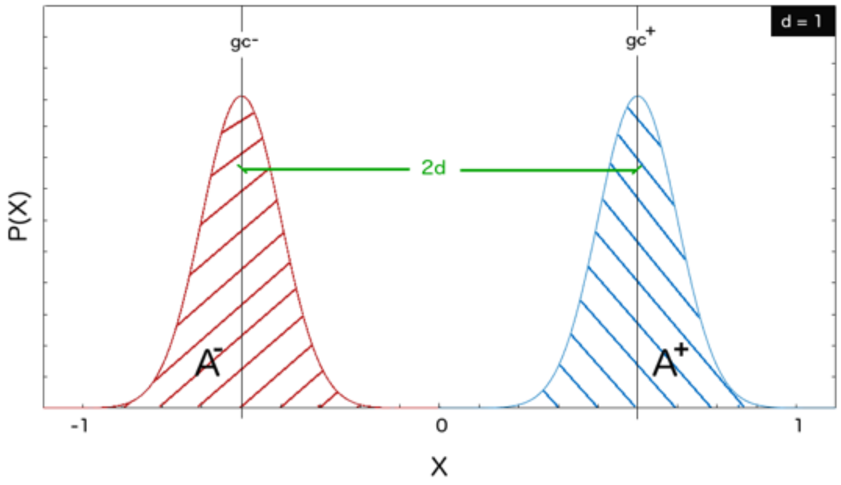


Figure 6: Network fragmentation metric: segregation as bimodal distribution of preferences

When there is no fragmentation *p*(X) is a unimodal distribution centered on zero, otherwise μ reaches its maximum when the distribution function is given by two Dirac deltas centered at -1 and +1 respectively.

The method we use to compute opinion fragmentation requires the distribution of the Elite users' opinion to be known as initial condition. Following Wu et al. (2011) in order to be classified as an elite, an user has to satisfy two conditions:

1. The volume of interactions: the ratio between the number of elites' interactions (such as a retweet between two different users) with other participants and the

total number of interactions occurring in the network must be at least equal to 40%.

2. The number of followers: higher than 100.000.

After classifying users, an analysis is needed to discover elites' opinion bias. In this study, two coders identified elites' political views by independently reading their tweets and collecting information from other sources and media when necessary. When elite users were affiliated with other media, such as in the case of TV networks or newspapers journalists or political commentators, we assessed the level of bias towards either candidate based on existing perceptions about the medium prevalent political orientation. We matched this with the results obtained from the web site http://mediabiasfactcheck.com that provides an assessment of the political bias of the most popular US media. By triangulating these different sources, we assigned the initial opinion to all the elite users included in our sample equal to +1 or -1.

Data-mining

We use data-mining techniques to identify a model to investigate the relative importance of four types of variables in explaining variations in online polarization: *topology of the conversational network determined by the reply structure, conversational metrics based on the keyword network, sentiment,* and *traffic*. We refer to the CRISP_DM model that defines the steps to mine a dataset and use R and Rattle as tools (Shearer et al., 2000) and use four different types of predictive models: *linear regression, artificial neural networks, decision trees,* and *random forests*.

Results

The proposed methodology was applied to Twitter conversations generated around the three television debates between Hillary Clinton and Donald Trump during the 2016 US elections. We selected the analysis of the backchanneling conversations related to the Presidential television debates because these events possess several desirable characteristics for our analysis. First, opinion fragmentation is favored by the bipolar characteristic of the US electoral system centered around two major parties. Second, television debates are highly popular, intensely advertised, and able to attract a broad audience. Consequently, these TV events trigger real-time backchanneling conversations with a significant amount of online buzz. Finally, Presidential debates are based on strict rules, in order to provide equal opportunities for each candidate and to share with the public her or his position concerning the issues being discussed.

Each of the three debates ran from 21:00 to 22:30 EST. The data collected covered the entire duration of the debate. After data cleaning and pre-processing, our dataset was composed of 71,835 tweets. Each debate was divided into time buckets of variable duration depending on the topic discussed by candidates during the event. Thus, buckets size is based on the moderator's questions. Following the methodology proposed by (Lipizzi et al., 2016), the streams were processed, and a database containing users, text, timestamps, and web analytics were created (tab.5).

Table 5: Spreadsheet example

SENDER	TIMESTAMP	TEXT	KEYWORD LIST
ThisJustIn_2	1474937017	Are Trump and Hillary having a rap battle…	Trump, Hillary, Rap, Battle
katiedukewits	1474937017	RT:@DonaldJTrumpJr:Backstage at #debatenight…	Backstage
rphawg3150	1474937017	RT:@igorvolsky:@fakemattingram…	no keywords

Across the three events, we defined 20 buckets, based on debate transcripts and the sequence of topics being discussed as defined in the moderator agenda: 7 for the 1st debate, 8 for the 2nd debate, and 5 for the 3rd debate (tab.6).

Table 6: Spreadsheet example

	Bucket	Period	Timestamp	Debate Topics
1st Debate	1	before	26/9/16(20.43)	Achieving Prosperity America's Direction Securing America
	2	during	26/9/16(21.00)	
	3	during	26/9/16(21.40)	
	4	during	26/9/16(22.03)	
	5	after	26/9/16(22.36)	
	6	after	26/9/16(23.29)	
	7	after	27/9/16(00.22)	
2nd Debate	8	before	9/10/16(19.59)	Future Generation/Terrorism Immigration/Taxes Energy Policy
	9	before	9/10/16(20.30)	
	10	during	9/10/16(21.00)	
	11	during	9/10/16(21.34)	
	12	during	9/10/16(22.10)	
	13	after	9/10/16(22.35)	
	14	after	9/10/16(23.24)	
	15	after	10/10/16(00.12)	
3rd Debate	16	before	19/10/16(20.09)	Supreme Court Economy/Fitness to be President Foreign hotspots/National Debt
	17	during	19/10/16(20.00)	
	18	during	19/10/16(21.34)	
	19	during	19/10/16(22.08)	
	20	after	19/10/16(22.34)	

In this phase, we extracted the network of retweets for each bucket and then displayed the networks through Gephi. Fig.7 shows the reply network extracted from the 6th bucket at the beginning of the first debate. The example shows strong segregation among the users of opposing candidates, reflecting that conversational interaction is based on ideological similarity and that the two antagonist groups do not interact significantly.

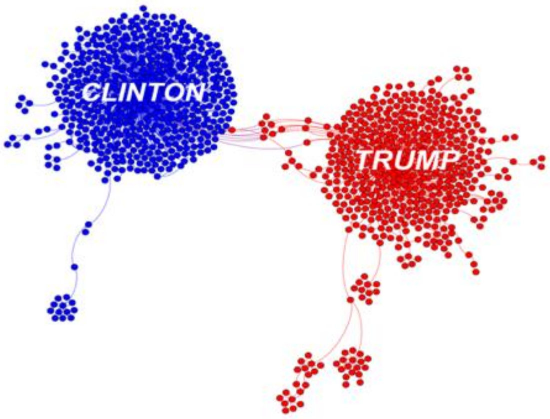


Figure 7: Reply network

Note: nodes represent users and edges represent retweets, while colors highlight different clusters.

Our scripts detected automatically potential elite nodes based on their number of followers and retweet received. The political bias of each elite was determined through web searches and tweets, as illustrated in the methodology section. Following that procedure, we identified 129 elite nodes, of which 68 leaning republican and 61 democrats. The disagreement between coders was resolved by additional search and discussion.

Tab.7 reports the results determined through the electric dipole metric in terms of population size, ideological distance, and overall fragmentation for the 1st debate.

Fig.8 shows the level of conversational fragmentation in time determined across the debates. The end of each debate is indicated through a dashed line. The higher this level is, the more participants tend to discuss and interact within their group. We also report the three most recurrent hashtags, positive, and negative words.

Table 7: Results for the first debate

BUCKET	DIFFERENCE IN POPULATION SIZE (ΔA)	IDEOLIGICAL DISTANCE (d)	NETWORK FRAGMENTATION (μ)
1	0,2164	0,9611	0.753
2	0,4420	0,9903	0.553
3	0,5780	0,9961	0.420
4	0,4402	0,9787	0.548
5	0,4176	0,9788	0.570
6	0,1421	0,9506	0.815
7	0,1448	0,9458	0.809

The interpolation in fig.3 shows that the average level of fragmentation increased as the campaign progressed towards election day. An Anova run across the three debates showed that level of fragmentation were different during the three events (F = 8.92, $p < 0.05$); post hoc analysis shows that the first debate differs significantly with both the second and the third ($p < 0.05$), while there is no significant difference between the second and the third. The first debate focused predominantly on the job and fiscal policy, and it was the one in which we recorded the highest cross-group interaction.

Within the first and the third debates, polarization roughly followed a U-shaped curve: the initial value decreased during the debate and then bounced back at the end, achieving a final value that was similar to or higher than the initial level. In the second debate, polarization had a more fluctuating behavior but did not decrease.

These trends reflect the dynamic predicted by our hypotheses: conversational interaction across the two opposing factions generally increased during the debate, and participants were exposed to opinions generated by users having different political leaning; eventually, inter-group interaction came to a halt and users got back to reply mostly to like-minded nodes. As a result, the overall fragmentation in the network ended up being the same as or higher than the initial value, a possible indication that users' attitudes towards their preferred candidate or party became even stronger after the online discussion.

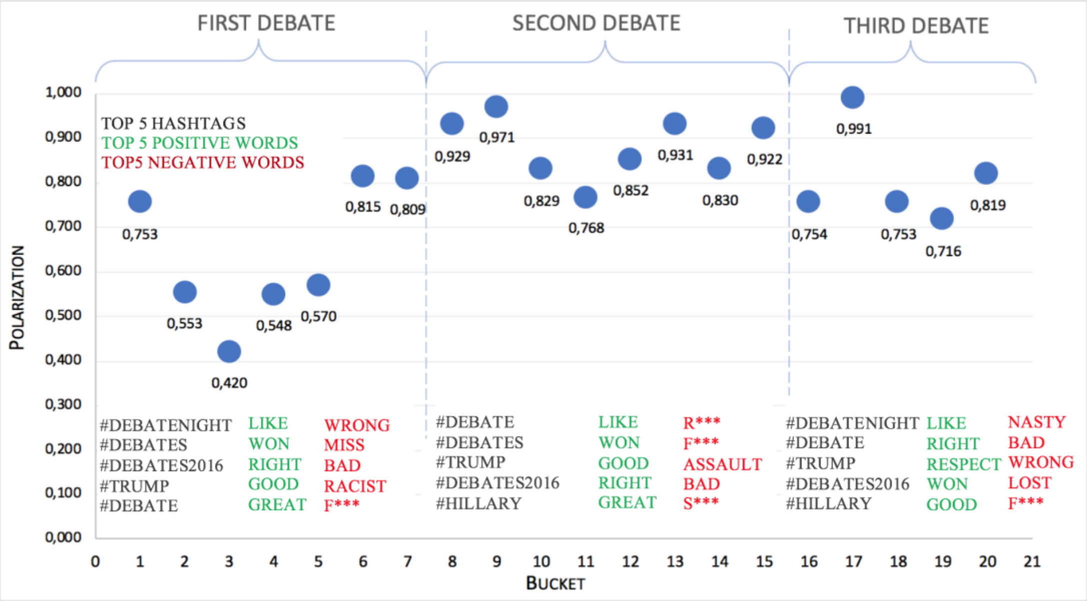


Figure 8: Visualization of network fragmentation measurements

In fig.9 we report the evolution of ideological views of the participants as it occurred during the debates in terms of preference distribution $p(X)$ computed over the two candidates (-1 Trump,+1 Clinton) in each time slice and in which the color indicates the number of participants to the online conversation.

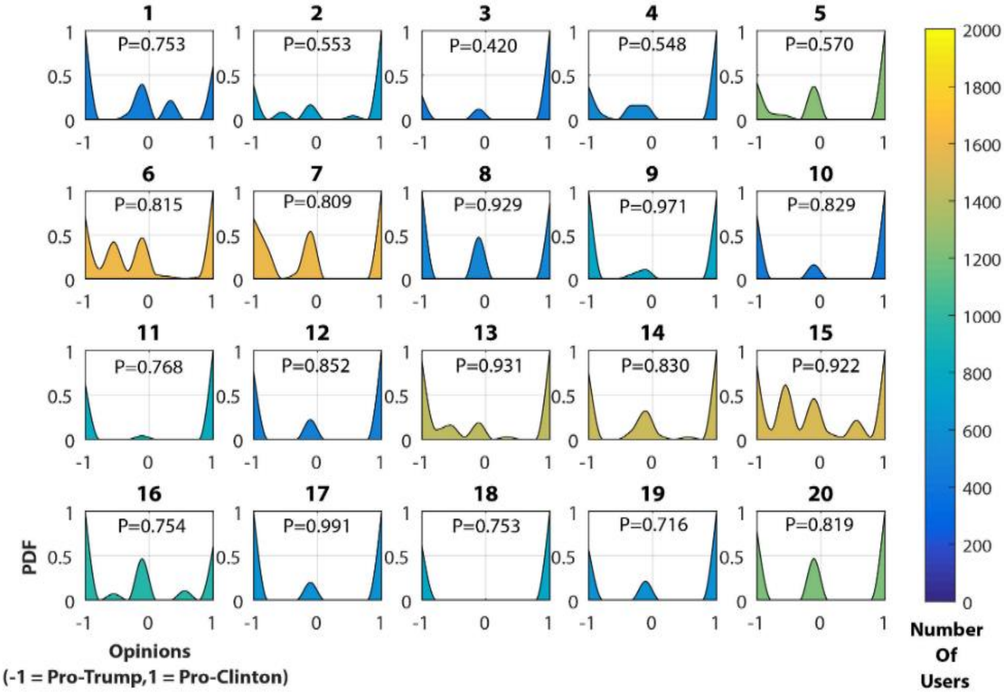


Figure 9: Overview of opinion distributions during the three debate

For the data-mining analysis, we consider 15 independent variables divided into four macro categories as specified in Tab.8:

1. Social Network metrics extracted from the reply network.
2. One traffic metric based on the number of tweets in a given bucket.
3. Conversational metrics, i.e., network metrics applied to the network of keywords extracted from the Twitter streams
4. Sentiment metrics obtained with Sentistrength.

Table 8: Evaluation table

CLASS	METRIC
SOCIAL NETWORK	Clustering coefficient Highest centrality Number of nodes Number of edges Density
TRAFFIC	Number of tweets
CONVERSATIONAL	Clustering coefficient Highest centrality Number of nodes Number of edges Density Words diversity [1] Sender diversity [2]
SENTIMENT[2]	Average sentiment Delta sentiment

Note: Conversational and Sentiment variables are measured on keyword network, while Social Network variables on RT's Network

[1] Measures within the considered bucket (Lipizzi et al., 2016)

[2] Measures obtained using Sentistrength (Thelwall et al., 2013)

We ran a preliminary exploratory analysis to detect correlation among our variables using Pearson's correlation coefficient. We found some positive correlation between fragmentation and sentiment (0.35/0.4) and some negative correlation between fragmentation and social network clustering coefficient (-0.5). We interpret the first correlation as a signal of the emergence of antagonism caused by the adoption of negative language by users (Ben-David and Matamoros-Fernandez, 2016). Regarding the second correlation, the clustering coefficient is a measure of the degree in which the nodes of a graph tend to be connected to each other; in this case, conversational fragmentation might be inversely related to the clustering coefficient due to the low number of edges between the nodes of different communities.

We applied Larose's supervised learning method (Larose, 2005) and used four different machine learning algorithms to model conversational fragmentation: decision tree, random forest, linear regression, and neural networks. We defined 70% of data as the training dataset and the remaining 30% as the testing dataset. As input variables for predictive models, we considered all possible combinations of the four types of variables. With the help of two Python modules (statsmodel and scikitlearn), which provide classes and functions for the estimation features of different predictive models, we evaluated the performance of all tested algorithms, both with traditional indices (such as the mean square error) and Bayesian indices (AIC and BIC).

Tab.9 summarizes the results. The minimum mean squared error (MSE) has been observed using content variables (0.020) or also through a combination of content and traffic variables (0.020), while the worst performance regards the combination of sentiment variables with traffic ones (0.43). When we combined the contributions of all four groups of variables, the error increased to (0.029), slightly less than the value obtained using only social network variables (0.031).

Table 9: Evaluation table

VARIABLES				BEST	BEST
C	N	S	T	MSE	ALGORITHM
X				0.020	DT
	X			0.031	RF
		X		0.038	RF
			X	0.027	DT
X	X			0.028	RF
X		X		0.030	RF
X			X	0.020	DT
	X	X		0.032	RF
	X		X	0.033	RF
		X	X	0.042	RF
X	X	X		0.030	RF
X		X	X	0.031	RF
X	X		X	0.029	DT
	X	X	X	0.034	RF
X	X	X	X	0.029	RF

Note:C=Conversational; N=Network; S=Sentiment; T=Traffic; DT=Decision Tree; RF=Random Forest

All the results show a strong consistency in indicating decision tree and the random forest algorithms as better models (see tab.10).

Table 10: Data-mining models comparison

DECISION TREE						RANDOM FOREST					
AIC		BIC		MSE		AIC		BIC		MSE	
TC	-4.388	TC	-4.596	C	0.020	C	-3.047	C	-3.255	C	0.027
C	-4.388	C	-4.596	TC	0.020	NC	-2.931	NC	-3.139	NTC	0.028
T	-2.945	T	-3.153	T	0.027	NTC	-2.902	NTC	-3.110	NC	0.029
NC	-2.052	NC	-2.260	NTC	0.030	N	-2.874	N	-3.082	TC	0.029
NTC	-2.052	NTC	-2.260	NC	0.030	TC	-2.686	TC	-2.895	NTSC	0.030
NT	-1.440	NT	-1.648	S	0.048	NTSC	-2.611	NTSC	-2.820	NSC	0.030
N	-1.440	N	-1.648	TS	0.048	SC	-2.557	SC	-2.765	SC	0.031
NTS	0.799	NTS	0.591	NS	0.048	NSC	-2.534	NSC	-2.742	N	0.031
S	0.799	S	0.591	NTS	0.048	NS	-2.400	NS	-2.608	TSC	0.032
TS	0.799	TS	0.591	SC	0.049	TSC	-2.325	TSC	-2.533	NS	0.033
NS	0.799	NS	0.591	NSC	0.049	NT	-2.031	NT	-2.240	NT	0.034
NTSC	0.860	NTSC	0.652	TSC	0.049	NTS	-2.023	NTS	-2.231	NTS	0.034
NSC	0.860	NSC	0.652	NTSC	0.049	S	-1.263	S	-1.471	T	0.038
TSC	0.860	TSC	0.652	N	0.049	T	-0.989	T	-1.197	S	0.039
SC	0.860	SC	0.652	NT	0.049	TS	-0.638	TS	-0.846	TS	0.043
GENERAL LINEAR MODEL						ARTIFICIAL NEURAL NET					
AIC		BIC		MSE		AIC		BIC		MSE	
S	-0.760	S	-0.968	S	0.055	S	3.528	S	3.320	S	0.076
SC	1.727	SC	1.519	T	0.085	N	6.238	N	6.030	N	0.120
T	4.118	T	3.910	SC	0.096	NT	6.449	NT	6.241	NT	0.123
N	6.238	N	6.030	N	0.120	TS	6.825	TS	6.616	TS	0.146
NT	6.449	NT	6.241	NT	0.123	T	7.910	T	7.702	T	0.163
TS	6.825	TS	6.616	TS	0.146	NS	8.183	NS	7.975	NS	0.173
NSC	7.459	NSC	7.251	C	0.148	SC	9.523	SC	9.314	NTS	0.232
C	7.536	C	7.328	NS	0.173	NTS	9.922	NTS	9.714	SC	0.294
NS	8.183	NS	7.975	TC	0.199	NSC	11.597	NSC	11.388	NSC	0.422
TC	9.246	TC	9.038	NSC	0.219	NTSC	14.678	NTSC	14.470	C	1.015
NTS	11.190	NTS	10.982	NTS	0.291	TSC	14.762	TSC	14.554	TC	1.374
NTC	11.571	NTC	11.363	TSC	0.611	NC	14.825	NC	14.616	TSC	1.738
NTSC	12.39	NTSC	12.17	NC	1.85	NTC	15.21	NTC	14.99	NTSC	6.76
TSC	13.75	TSC	13.54	NTSC	22.43	C	15.49	C	15.29	NTC	134.96
NC	15.294	NC	15.085	NTC	40.061	TC	15.517	TC	15.309	NC	169.32

Note: S=Sentiment, N=Network, T=Traffic, C=Conversational

Discussion

Our empirical results provide evidence that conversational interaction across members of antagonist groups and, thus, exposure to diverse opinions do occur in online political debates taking place through Twitter backchanneling conversations. In all the three debates, this conversational interaction evolved in time with a similar trend (fig.8) by increasing during the debate compared to the interaction taking place before and after the debate itself. As predicted by common ground and persuasive argument theories, Twitter users do engage with diverse opinions. However, this interaction seems more dictated by the occurrence of a polarizing event (the debate) than by the need to access new information and possibly reconsider existing beliefs.

While Twitter allows users to interact with virtually any other participant, our data show that this openness can backfire, at least when it comes to political discussions: instead of favoring revision of pre-existing beliefs, interaction with other users provoked both antagonist segregation.

While concepts such as filter bubbles or echo chambers have proven problematic to define and identify, we think that tracking conversational interaction across ideologically segregated communities could be a promising perspective in the analysis of this phenomenon. The bubble metaphor can be applied to our results as well in a dynamic way since the fragmentation metric we adopt in this work reveals the inflating and deflating of ideologically homogenous retweet sub-networks (fig.7 and 9).

Therefore, the expectation that online conversations should favor exposure to diverse beliefs is supported by our data; however, the behavior following this exposure does not seem conducive to a more open and fertile debate. Not only the temporary increase is followed by an decrease in interaction across the two groups, but, upon closer inspection of the content of the tweets, we found out that the interaction among users belonging to antagonist factions is typically oriented at attacking opposite points of view or at reasserting prejudicial positions, as predicted by the persuasive argument theory. The decline of the dipole metric used in this study (fig.8) signals that, at least for some time, users do interact with participants holding opposite opinions, but the subsequent increase reveals that contributors eventually turn back to retweet like-minded peers' posts. In conclusion, our results suggest that online ideologically, segregation is the outcome of the combination of two factors: predominance of interaction with like-minded peers coupled with antagonist interaction with the opposite faction during polarizing events.

Through our analysis, we identified a model based on a combination of online traffic and conversational metrics that achieves satisfying performance in terms of MSE (0.020). This result shows that the actual content being exchanged plays a role in determining polarization more than the topology of the reply network, meaning that the content being exchanged is more critical than the who-talks-to whom interaction pattern. While the topology of the reply network provides the social highways along which diffusion of information can take place, and cohesive groups can be created and maintained, it is the content being exchanged that determines conversational interaction. A more in-depth analysis shows that conversations with higher lexical diversity are more likely to be associated with higher values of fragmentation: people belonging to different, antagonist groups use a more diverse lexicon when communicating with members of their group than when addressing members of the other party. Thus, intra-group conversations appear to be more articulated than intergroup ones, perhaps another sign that interaction among users bearing different opinions is less sophisticated and aimed at simplifying the message rather than to compare and assess diverse positions on different topics and issues. Similar results were obtained for other structural metrics extracted from the keywords network: conversations with like-minded peers appear to be more structured and cohesive as signaled by higher density, higher topic clustering indicator, and higher betweenness centrality among topics. We explain this result in terms of difficulty of accumulating common ground due to the combination of strong pre-existing attitudes and limited affordances provided by the medium to support more sophisticated discussion. It is also interesting to note that sentiment alone has a weak exploratory power.

While we show that online and offline political segregation share similar features, our study does not provide any evidence regarding whether online conversations and users end up being more polarized than in the face-to-face case. Whether the Internet makes online political debate worse, and specifically more polarized, is a very controversial question and scholars and media analysts tend to distribute across the whole spectrum connecting e-democracy enthusiasts with vocal skeptics. In order to answer this question, empirical panel studies designed to compare systematically online and offline debates are needed (Settle, 2019). However, based on our findings, we argue that SM, and Twitter, in particular, possess design features that might exacerbate the natural tendency of groups to ideologically and value-based segregate discussions. A safer statement would be that, at the very least, Twitter does not natively offer any countermeasure to this phenomenon and, more broadly, to other group thinking biases. First, based on the persuasive Argument theory, SM makes it easier to interact conversationally with users having an opposite opinion, which makes argumentative defenses of pre-existing attitudes and positions prominent over the rational assessment of alternative stances. Second, Twitter makes it very easy to establish preferential connections among like-minded users as well as to broadcast messages to this audience through its reply structure. Thus, the emergence of cohesive groups and ideological bubbles can be facilitated by these features. The combination of persuasive argumentation and selective exposure creates both the conditions for more intense confrontation and emergence of information cocoons that filter out content that is not aligned with the dominant preference. Sustained and prolonged antagonist interaction can nurture and empower the natural tendency of human beings to prefer homophilous social links and affiliation to groups that provide social identity. We think that, at the very least, the ability of SM to expand our conversational sphere could be the culprit behind the negative consequences of exceedingly partisan political discourse. It would be interesting to run controlled studies in which participants use a medium with less or more costly conversational features (e.g., Twitter with the retweet function turned off) and check which situation is more conducive to partisan behavior, judgment polarization, and ideological fragmentation. We predict that with less conversational interaction the quality of the debate would improve; however, we are aware that removing conversational affordances makes online interaction less socially meaningful and that, outside of labs and academic attempts, in the real world it is exactly the management of this tradeoff that is highly problematic (Iandoli et al., 2014).

This study has implications for other domains as well. If exposure to polarized information has behavioral consequences, inducing polarization in a debate could push participants to make (or not make) certain choices. For instance, it is intriguing that ideological attachment has received attention in marketing through studies on polarizing brands. Luo et al., 2013 evaluated various brands' polarization and examined the relationship between polarization and stock market returns, suggesting that highly divisive brands tend to perform more poorly than others, but they also tend to exhibit relatively little variation in stock price (Luo et al., 2013). As Guy Kawasaki says, companies should not be afraid of polarizing people, rather they should worry not to leave them indifferent (Kawasaki, 2004). Ideological segregation can offer numerous advantages, such as obtaining passionate responses from users, brand advocacy, differentiation, and an increase in sales (Alvaro et al., 2014). Given the responsiveness of online communities, it is crucial for a company to understand how these groups side in favor of or against a brand, and how these antagonist communities interact (He et al., 2013).

The analogies between online political debate and consumers' behavior could offer fertile ideas for new research. In both contexts, there is the presence of a community (consumers or voters) to whom the promotional messages are addressed (advertising or electoral propaganda) that aim at soliciting a given course of action (purchase or vote). More than other media, the Internet is probably contributing to both the politicization of marketing through brand advocacy and the marketization of politics. In both domains, it is becoming imperative to listen to what the target audience likes and wants, to track this listening using analytics, and to respond appropriately to create appreciation and consensus. The obsession towards the listening aspect can potentially stifle innovation in both the creation of new political perspectives and vision and in new product development. Political marketing can offer essential insights to understand the phenomenon of polarization in promotional contexts, and this could be a promising perspective for research in these areas.

Conclusion

In this chapter, we have presented an empirical analysis of conversational interaction enabled by SM in online political debate and of its determinants using data extracted by Twitter backchanneling conversations. Our analysis offers evidence that online ideological segregation based on political difference exhibits similar dynamics to offline political debate; however, we also show that sustained and protracted conversational confrontation can be facilitated by the affordance provided by SM making the "enemy" a frenemy that always available and always accessible (Settle, 2019).

This study has some limitations. First, we do not systematically compare online and offline groups, so we cannot ascertain whether ideological segregation is more intense in online or face to face settings. A second significant limitation resides in the data-collection procedure: the free API for tweets download provides only a sample of all the messages. The alternative is to connect to the full Twitter pipeline, but this connection is rather expensive. The free method is, however, still capable of capturing a large number of messages randomly in a short time window, assuming the event is popular enough to generate sufficient online buzz. Third, our analysis is exploratory, and further work is required to replicate the study on a larger dataset as well as to move from the exploratory level to the development of a causal model linking polarization to its determinants. Fourth our metric does not measure content polarization, i.e., the formation and sharing of more extreme opinions on specific topics due to group bias and mutual reinforcement. However, only the extent to which SM users engage with users has opposite positions on preferences. Finally, our method requires the identification of elite users with known preferences towards either alternative; this requirement may not always be easy or possible to satisfy in other applications.

Future developments could be to perform a more in-depth investigation of which conversational features and associated variables play a role in online ideological segregation. If contents and conversational dynamics have a significant impact, framing a message could be more impactful than infiltrating a social network to disrupt the existing connections, which will be achieved as the consequence of a valid "content first" strategy.

Finally, a big open question regards the effect of online ideological segregation and opinion polarization on actual behavior. As outlined above, a positive impact could help to be applied in several domains, such as online advertising, health behavior, and political campaigning. If empirical evidence should show that these consequences are exaggerated, the concern of the cyber skeptics could be downsized along with some proposed interventions aimed at reducing communication freedom on the Internet.

CHAPTER
THREE

RADICAL SHIFTS: THE IMPACT OF IDEOLOGICAL DIVERSITY IN ONLINE DISCUSSION ON OPINION RADICALIZATION AND INTENTION TO MOBILIZE

Summary

Online discussions are often accused of fomenting GP. Although online discussions are ubiquitous, they have received less attention in studies on online GP. In particular, no systematic evidence is available to show whether polarization induced by online discussions can transfer to intention to mobilize to support a cause. Besides, since GP can be the result of exposure to both similar and adversarial opinions, we intend to compare systematically which of these two conditions is more polarizing.

To address these research gaps, in this chapter, we present a study in which subjects were assigned to online discussion groups characterized by different levels of ideological diversity. Our data show that subjects belonging to groups with diverse participants became more radical in terms of attitude and subjective norms, while participants to like-minded groups reinforced both subjective norms and perceived behavioral control but did not become more extreme in their attitude. In none of the two conditions, polarization transferred from opinions to intention to mobilize. Overall our results provide evidence that online discussions can actually make participants more extreme in some respect and that GP dynamics are likely to fulfill a different social function in ideologically homogeneous VS heterogeneous groups.

Introduction

As shown by the literature review (see chapter one), in his essay titled "Why They Hate Us," written in the aftermath of the 9/11 terrorist attack, Cass Sunstein (2002) accused the Internet of having evolved into a discursive space nurturing radicalization and political extremism. More specifically, Sunstein proposed that online discussions occurring among ideologically biased, like-minded peers favor GP via the creation of biased 'echo-chambers' (Sunstein, 2007, 2002a, 2001). Subsequent research confirmed this view by providing evidence of GP in online groups not necessarily engaged in extremist political activities (Grömping, 2014; Messing and Westwood, 2014; Stroud, 2010).

More recent studies show that while echo-chambers have been abundantly studied in ad hoc experimental designs, their importance in the current online landscape might have been overstated (Balcells and Padró-Solanet, 2016; Bright, 2018; Shore et al., 2018; Weber et al., 2013). The Internet facilitates unprecedented exposure to alternative points of view by favoring access to diverse information through weak ties (Bakshy et al., 2015; Barberá, 2014; Grabowicz et al., 2012) and by allowing users to create a media diet that is quite varied and procured through serendipitous search (Dubois and Blank, 2018; Flaxman et al., 2016).

Access to diverse information, however, seems not able to guarantee a more balanced and rational debate in online discussions. By making antagonist points of view more visible and by favoring involuntary exposure to ideologically charged content (e.g., political messages), the Internet can increase polarization due to the participants' need to preserve their social identity (Settle, 2019; Yardi and Boyd, 2010).

Thus, existing research identifies both in lack of diversity and exposure to diverse conflicting points of view as viable mechanisms for the emergence of polarization in online discussions.

In this chapter, we intend to contribute to the study of the polarizing effects of online discussions in three ways. First, we compare polarization in online discussions among like-minded peers VS the more common situation in which participants enter into conversational engagements with 'disagreeable others' (Settle, 2019), i.e., people bearing antagonist points of view. Second, we want to ascertain whether opinion polarization transfers to an increase in the intention to mobilize to support the endorsed position; the presence of a transfer effect from words to action can significantly magnify the concerns about the consequences of online polarization. Third, compared to other political behaviors people report in online media, conversations on sensitive topics have received much less attention in the literature (Settle, 2019), so with this study, we intend to complement existing research that has focused primarily on exposure to ideologically contrary information (Asker and Dinas, 2019; Bail et al., 2018) as opposed to active conversational engagement.

We use Ajzen's Theory of Planned Behavior (TPB, Ajzen, 1991) to observe whether changes occur at the intention level and/or at the level of its antecedents (attitude, subjective norms, or perceived behavioral control) and offer a theoretical framework in which social identity is complemented by a theory of argumentation-based polarization (Mercier and Sperber, 2011).

The next section presents the theoretical framework and the research hypotheses relating to participation in online discussions on opinion polarization and intention to act. Then, we present an experiment conducted with 356 participants divided into small groups (average size 17 members) and assigned to either one condition (like-minded VS cross-minded participants) to compare the effects of participation to an online discussion about a divisive issue on belief and intention formation.

Our results show that subjects belonging to groups with diverse participants became more radical after the discussion in terms of attitude and subjective norms, whereas participants in the like-minded peer situation reinforced both subjective norms and perceived behavioral control but did not become more extreme in their attitude. Finally, we report that a significant shift in the intention to mobilize for the cause did not occur in any of the two conditions, while members of both cross-minded and like-minded groups became more radical compared to subjects who did not actively participate in the discussion.

Theoretical framework and research hypotheses

In this paper, we adopt the TPB to measure whether active participation in online discussions on controversial topics can determine opinion and intention shifts. More specifically, we intend to assess these shifts in function of the level of ideological diversity in a group (like-minded VS cross-minded participants).

According to TPB, intention formation has three precursors: attitude, subjective norms, and perceived behavioral control reflecting, respectively: i) the predisposition to act positively or negatively toward some object, ii) the perceived social desirability of undertaking the behavior; and iii) the perception of how feasible it is to perform the behavior. Overall, we hypothesize that active participation in online discussion will be responsible for opinion and intention polarization, and in the following, we provide a detailed explanation of the theoretical mechanisms determining such shifts. We adopt the following definitions:

- Online GP is the tendency of individuals to become more extreme in their opinions on a controversial issue as induced through a mix of group dynamics created through online interaction and exposure to biased information.
- A controversial issue is a contentious subject on which people tend to take opposite positions typically due to irreconcilable differences at the political, cultural, or ethical levels.
- A group of like-minded individuals is a group with a substantial majority of individuals having the same position on a controversial issue.
- A group of cross-minded individuals is a group with a balanced composition of individuals having opposite positions on a controversial issue.

With opinion (intention) polarization, we refer to individual opinion (intention) shifts towards a more extreme position than the one initially held by the individual.

The proponents of the echo-chamber theory (Sunstein, 2002a) justify opinion polarization via online discussion through two fundamental mechanisms: impaired information processing via exposure to biased information and social pressure to conform. First, discussion among like-minded peers reduces and suppresses information diversity because of the similarity between participants' preferences and background and of the common knowledge effect (Gigone and Hastie, 1993). Second, discussions favor self-confirmation bias (Bessi et al., 2016a; Cho et al., 2016; Sunstein et al., 2017; Yardi and Boyd, 2010) and reinforce homophily (Bessi et al., 2016a; Stroud, 2010).

In addition to faulty information processing, online discussions among like-minded participants favor pressure to conform to the group norms and dominant beliefs (Asch, 1963), satisfy the need for affiliation and social influence (Sunstein, 2002a), and can enforce inhibitory mechanism such as the so-called spiral of silence (Hampton et al., 2014; Noelle-Neumann, 1974).

Therefore, we hypothesize that:

> *Hp 1.1 Active participation in online discussions on a controversial issue among like-minded participants favors opinion polarization.*

On the other hand, contrary to the expectations that group deliberation will produce more moderate options, the exposure to diverse opinions can produce polarization as well (Balcells and Padró-Solanet, 2016; Bright, 2018; Dubois and Blank, 2018; Shore et al., 2018; Yardi and Boyd, 2010). Social identity theory (Tajfel, 1982) accounts for this phenomenon by positing that polarization emerges from the need to preserve the group identity against the opposite faction by nurturing an "us VS them" rhetoric driven by the process of social categorization. Following this theory, Settle (2019) argues that on SM such as Facebook online polarization can be significantly enhanced by the visibility of antagonist points of view through the exposure to potentially divisive information, the access to News injected into individual newsfeed by polarized elite and media, and the participation in online Discussions (END framework). More importantly, this enlarged access to disagreeable others induces affective polarization (Iyengar et al., 2019), i.e., a form of polarization that transfers from the disliking of content and ideas to the disparagement of individuals that have opposite opinions. The END framework thus predicts that participation in online discussions will favor the production of explicit content that makes social categorization easier and the emergence of affective polarization more likely.

Finally, not only online discussions can expose to controversial content, but the production of this content occurs through polarizing conversational mechanisms. Mercier and Sperber (2011) offer an elegant argumentation-based theory of reasoning that is specific about the type of conversational mechanism that can reinforce social identity: the need for evaluating and produce arguments to persuade others. Under this need, the participants to a discussion will build arguments to support positions that are consistent with their pre-formed opinions and attack positions they do not agree with. Thus, when exposed to opinions they do not like, "[participants] reflect less on the item itself than on how to defend their initial attitude" (Mercier and Sperber, 2011, p. 67).

We then hypothesize that:

Hp 1.2 Active participation in online discussions on a controversial issue among cross-minded participants will favor opinion polarization.

In addition to the impact on opinion polarization, we also want to test whether participation in online discussions has an impact on intention formation. More specifically, with reference to the condition of like-minded individuals, we hypothesize that participation in online discussions on a controversial issue will affect intention formation antecedents and then intention itself to mobilize to support a cause, as predicted by the TPB. In this way, we aim at testing empirically Sunstein’s theory of radicalizing echo-cambers (2002) that describes an escalation process in which online conversations among like-minded peers help to reinforce the dominant beliefs to the point of pushing (some) participants to radicalize and mobilize in support of the cause. The following hypotheses are derived directly from Ajzen’s model. More specifically, we hypothesize that active participation in an online discussion on a controversial issue among like-minded participants will:

Hp 3.1 increase the degree to which mobilizing in support of the cause is positively valued (attitude)

Hp 4.1 increase the degree to which mobilizing in support of the cause is perceived as socially acceptable (subjective norms)

Hp 5.1 increase the degree to which mobilizing in support of the cause is perceived as more feasible (perceived behavioral control)

Following TPB, the effect of such participation on intention determinants will imply a shift also at the intention level:

Hp 2.1 Participation in online discussions on a controversial issue among like-minded peers will increase individual intention to mobilize.

Hp 3.1 can be justified in several ways. First, in the absence of controversy, participants are more likely to accept that their point of view as valid (Warner, 2010). Second, lack of diversity favors confirmation bias, which is responsible for the selection of biased narratives and facts as well as for the discrediting of the information that can harm the dominant point of view (Baum and Groeling, 2008; Suhay et al., 2018; Yang et al., 2016). Consensus and the availability of biased information will make appear as righteous and appropriate to mobilize to support the cause.

Hp 4.1 tests for the effect of participation in biased online discussions on the perception of how socially desirable might be to undertake a specific behavior. Prior studies demonstrate that people conform to the opinion of other group members and converge to social norms because of their need to feel accepted and be connected by others within a group (Festinger, 1954). Membership to certain groups helps to identify boundaries within which it is easier to discriminate between appropriate or inappropriate conduct (Postmes et al., 2000). Within these boundaries, people feel a higher motivation to comply because of a desire to maintain their reputation in the group (Sunstein et al., 2017). Therefore, we posit that the discussion with like-minded peers exalts the perception that is undertaking action to support what the group values as desirable is socially appreciated.

Finally, concerning the effect of biased online discussions on perceived behavioral control (Hp 5.1), we expect that a group of like-minded peers will provide an environment in which people can expect to find resources and help to take action. Increasingly, people rely on online tools to perform "how-to" searches or look for direct help to carry out tasks (Livingstone et al., 2005), by taking advantage of supportive communities providing free advice and resources for action (Rainie and Wellman, 2012). When it comes to social activism and mobilization, Internet and SM have become essential organizing tools, as in the case of the Arab Spring, the MeToo, or the Occupy Wall Street movement (Bruns et al., 2013; Ince et al., 2017; Kuo, 2018; Tremayne, 2014). Since supportive and ideologically friendly online groups can offer multiple types of help and support, group members will perceive it easier to mobilize to support the cause. In all the above examples, online conversations occurring in friendly and ideologically cohesive communities, are the most critical interaction affordance.

Relying once again on TPB, we formulate specular hypotheses relating active participation in online discussions also for groups composed of cross-minded individuals. More specifically, we hypothesize that active participation in online discussions on a controversial issue in cross minded groups will:

Hp 3.2 increase the degree to which mobilizing in support of the cause is positively valued (attitude)

Hp 4.2 increase the degree to which mobilizing in support of the cause is perceived as socially acceptable (subjective norms)

Hp 5.2 not increase the degree to which mobilizing in support of the cause is perceived as more feasible (perceived behavioral control)

And that:

Hp 2.2 Participation in online discussions among cross-minded individuals on a controversial issue will increase individual intention to mobilize.

Hp 3.2 and 4.2 can be justified in a similar way as in the case of like-minded peers. Reinforcement mechanisms will be available to help both attitude and social norms to become more intense. Although in this case, the reinforcement will work to support members of a faction to differentiate their identity and defend their positions against the opposite faction, as predicted by social identity theory.

Regarding Perceived Behavioral Control (Hp 5.2), we hypothesize that groups of cross-minded participants will find less help and support because of the adversarial climate of the discussion. First, if the primary objective of a faction is to attack and possibly defeat the antagonist group rhetorically, there will be less time and energy to dedicate to mutual help and information seeking. Second, exposure to the attacks and the strength of the antagonist faction will not facilitate confidence building.

Methodology

Subjects

Participants in this study included students recruited from a large High School located in Italy. The subjects were aged between 14 and 20 years[3]. Students were quite evenly distributed in terms of age and gender (56% male). Participation in the study was part of optional, credit-bearing activities.

Procedure

Figure 10 reports all the steps of the study protocol. In the first step, participants were asked to fill a questionnaire to express their level of interest and ideological standing on four topics presented in the form of a moral dilemma ("Do you think X is right or wrong"). The choice of the four topics for discussion was based on the following criteria: i) the topic had to be controversial, and either ideological standing should lead to alternative, incompatible choices, ii) the topic had been current and exciting for most of the subjects, and iii) participation to the discussion should not require the possession of specialized knowledge or skills.

Through a series of meetings with teachers, school administrations, student representatives, and fellow researchers, four topics were identified as possible candidates: universal income, mandatory vaccination, legalization of "soft drugs," and adoption of more stringent urban mobility regulation to reduce pollution and congestion.[4]. Subjects

[3] The research protocol was approved by the IRB of the University. Underage participants had to provide a written authorization signed by their parents or guardians to be admitted to the activities. The study was part of a School promoting awareness and knowledge in the use of social media. The project procedure was discussed among a group of teachers who were in charge of the project and approved by the school principal and board. The choice of the discussion topic was carried out together with teachers ensuring both consistency with the project and a safe discursive space for the students.

[4] It is important to note that by defining these topics as controversial, by no means, we imply that they are such according to science, but only that these topics tend to be divisive among laypersons and non-experts. For instance, the benefits of mass vaccination have been clearly proved by Science; however, a not negligible fraction of the general population has gathered in anti-vax movements for various reasons ranging from religious beliefs to libertarian ideology or distrust in official Science.

were then polled through a questionnaire to express their level of interest and opinion polarization on each topic.

In step 2, subjects were also asked to fill a second questionnaire aimed at measuring TPB constructs. Based on the answers to the survey, participants were assigned to discussion groups, with each group in either one treatment condition: groups composed by like-minded peers VS groups composed by cross mind-minded participants. Subjects were randomly assigned to groups, and subjects from the same class were allocated in different groups in order to simulate discussions with unknown participants or weak ties, a setting that has received little attention in the literature on online polarization (Settle, 2019).

In step 3, subjects participated in an online asynchronous discussion for one week. Finally, in the last step, TPB constructs and individual position on the issue were measured again after the discussion was over.

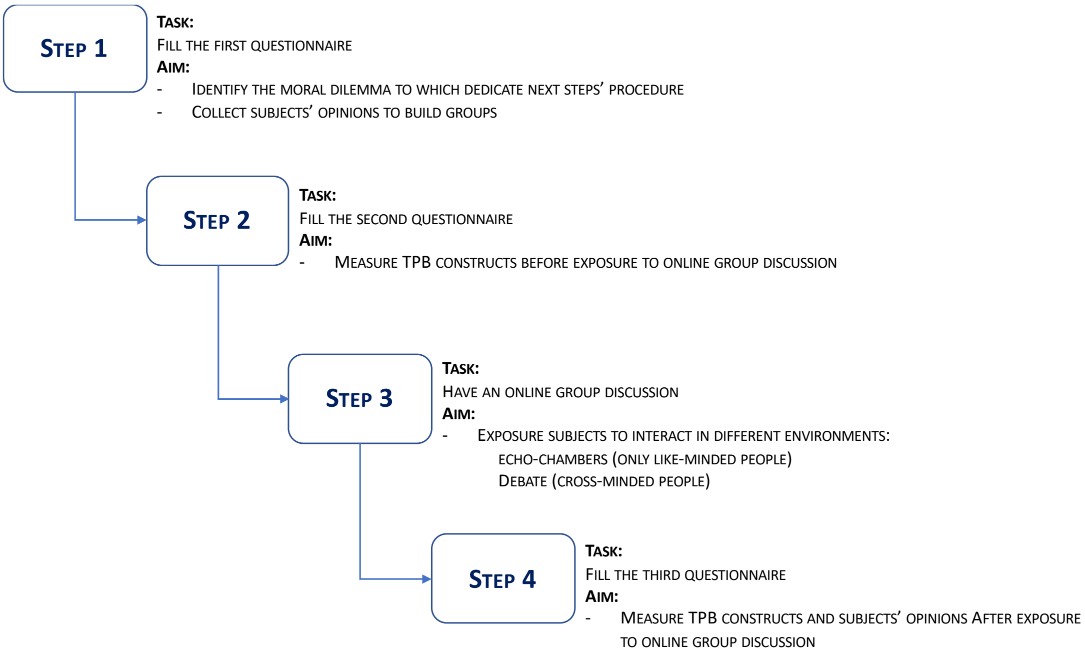

Step 1
Task:
Fill the first questionnaire
Aim:
- Identify the moral dilemma to which dedicate next steps' procedure
- Collect subjects' opinions to build groups
Step 2
Task:
Fill the second questionnaire
Aim:
- Measure TPB constructs before exposure to online group discussion
Step 3
Task:
Have an online group discussion
Aim:
- Exposure subjects to interact in different environments:
echo-chambers (only like-minded people)
Debate (cross-minded people)
Step 4
Task:
Fill the third questionnaire
Aim:
- Measure TPB constructs and subjects' opinions After exposure to online group discussion

Figure 10: The proposed procedure

Measurements

Through the first questionnaire[5], demographics data and opinions about the four selected topics were collected. Subjects were asked to rank topics from the most to the least interesting. Following Myers and Lamm (1976), subjects' position on an issue was collected through a set of 5 questions and measured on a five items Likert scale with a central neutral answer ("I do not know"), surrounded by four biased items on each side ("extremely favorable," "favorable," "unfavorable," and "extremely unfavorable").

The second questionnaire[6] was constructed following the guidelines provided by (Ajzen, 2004; Francis et al., 2004) for the measurements of TPB constructs. The questionnaire was divided into four sections, one for each Ajzen construct and included 30 questions of which 18 for indirect and 12 for direct measures[7]. For these questions, participants were asked to express their degree of agreement with a statement on a 5-point Likert scale (from "Strongly disagree" to " Strongly agree."

More specifically, each question in the second questionnaire aimed at measuring TPB constructs concerning the intention to participate in a public initiative to support the user's preferred choice. Following Francis et al. (2004), participants' attitude toward the behavior was assessed using three semantic differentials with respect to the statement "I think that participating to the initiative is: harmful/beneficial, useless/helpful, unsatisfying/satisfying). Responses were collected through a 5-point scales (e.g., extremely harmful (1), I do not know (3), extremely beneficial (5)).

To assess subjective norms, participants were asked the following questions: "If I participate to a public initiative to support the cause, people who are important to me would" (fully disapprove (1), fully approve (5)); "Most people who are important to me

[5] The questionnaire is available at this link translated in English (the original version was administered in Italian) https://forms.gle/Q86N63zFh9uzqcWF6

[6] https://forms.gle/EwqG1WKEFh9ipHLS6 (English translation)

[5] TPB constructs can be measured directly, e.g., by asking respondents about their overall attitude, or indirectly e.g., by asking respondents about specific beliefs and outcome evaluations. Direct and indirect measurement approaches make different assumptions about the underlying cognitive structures, and neither approach is perfect. According to Francis et al. (2004), it is preferable to be redundant and use both methods to increase robustness.

would think that participating to the initiative is (e.g., totally undesirable (1), totally desirable (5)).

Perceived control behavior, instead, was assessed through five items. Participants were asked questions like: "How much control do you have over whether you participate or not to the initiative?" (very little control (1), a great deal of control (5)); "For me to participate to the initiative is (very difficult (1), very easy (5))"; "If I wanted to, I could easily participate to the initiative" (strongly disagree (1), strongly agree (5)).

Finally, the behavioral intention was measured with items like: "I intend to participate (extremely unlikely (1), extremely likely (5)); "Will you plan to participate?" (definitely intend not to (1), definitely intend to (5)). An additional question was added at the end of the questionnaire through which participants were asked to indicate which kind of public initiative they would be more likely to support and participate to. The answers were ranked in terms of decreasing level of commitment: I intend to become an active member of an association dedicated to the cause (5), I intend to participate to a public walk (4), I intend to donate 10 euro (3), I intend to sign an online petition (2), and I do not intend to engage in any of the above (1).

With the third questionnaire, subjects' opinions and intentions were measured again after the discussion.

Opinion and intention construct scores were computed through a simple average of the associated items. Opinion and intention changes (shifts) were computed, subtracting the scores obtained after and before the discussion (post-pre).

Online discussion

After completing the second questionnaire, each subject received an email invitation with the link to a video providing necessary information on the tissue to discuss and of the rules of the discussions[8]. The discussion took place through an online forum in which the subjects could only interact with members of their group. The subjects participated in

[8] https://www.youtube.com/watch?v=hk5FqhuKXyM

a one-week asynchronous discussion. To ensure anonymity, users were allowed to pick a nickname.

Choice of the unit of analysis

A nested-ANOVA is proposed to ascertain whether the analysis of data should be based upon groups or individual observations (Kenny et al., 1998; Morran et al., 1990). Even when individuals are randomly assigned to groups and experimental conditions, it cannot be assumed that individual observations will remain independent of each other, given the interactive nature of the task (Morran et al., 1990). Previous studies on small group research have shown (Kenny et al., 2002, 1998; Kenny and Judd, 1986) that if the individual is used as the unit of analysis, the assumption of independent observation is likely to be violated when the task requires intra-group interaction because subjects may influence each other. Since ANOVA is not robust with respect to the violation of the independence assumption, analysis based on individual observations will likely produce inflated p-values.

Kenny, Kashy, and Bolger (1998) suggest a procedure to identify the appropriate choice between individual VS groups based on the presence of nonindependence in the data within the groups.

A group effect (nonindependence) occurs if individual scores within a group are more similar to one another than the scores of individuals in different groups. In order to assess and estimate the degree of nonindependence, the intraclass correlation for each metric of interest must be computed and tested for statistical significance (Kenny et al., 2002). A significant test reveals the presence of group effects, and the consequent need to use groups as unit of analysis. In order to deal with the likely presence of nonindependence, we designed our study by distributing participants in 20 groups with an average size of 17 members, ten in each condition, as opposed to using fewer large discussion groups.

Results

General overview

Three hundred fifty-six (N = 356) students completed all tasks included in the procedure, accounting for 65,08% of the students initially enrolled. The students that filled only the pre and post discussion questionnaires but did not participate in any online group discussion were included in a third group and labeled as Offline participants. Descriptive statistics and group composition by gender are reported in fig. 11, along with some statistics regarding the most used online SM in the cohort.

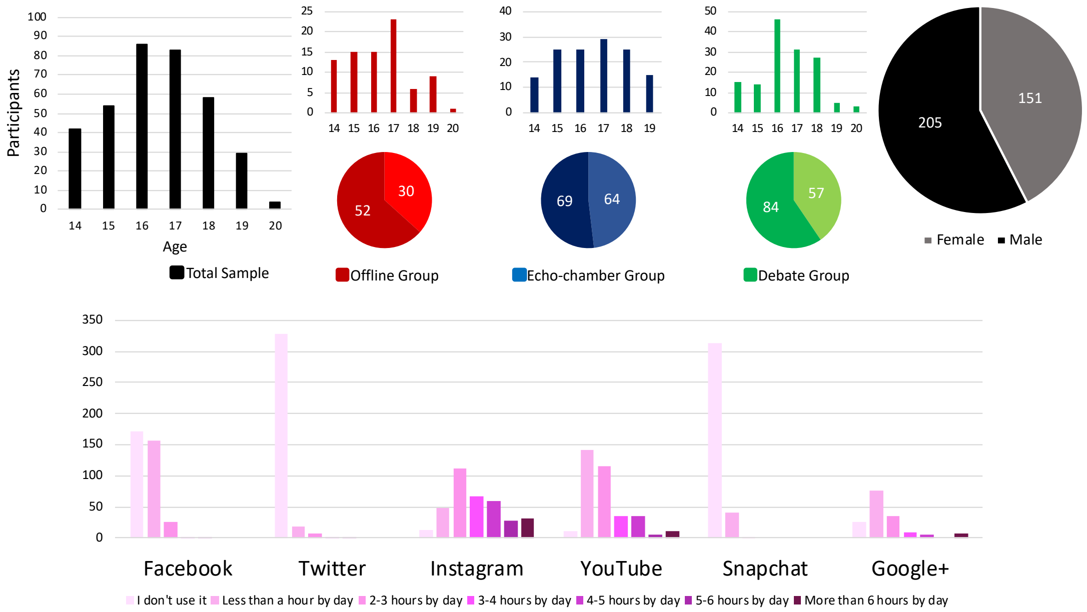


Figure 11:General overview

Questionnaires validation

Constructs used to measure TPB and opinion polarization included in the questionnaires were assessed in terms of reliability and validity.

As far as reliability is concerned, acceptable values of Cronbach's alpha were obtained for each construct (Tab. 11). The table also reports the results for the discriminant validity. As the square root of the average variance extracted (AVE) is much larger than the inter-construct correlation between any two constructs, the discriminant validity test is satisfactorily passed (Fornell and Larcker, 1981).

Table 11: Reliability and validity analysis

Construct	Item	Cronbach's Alpha	AVE	Sqrt(AVE)	Inter-construct correlations
P	5	0.815	0.426	0.652	-0.304; 0.059; -0.069; 0.133
A	3	0.882	0.726	0.852	-0.316; -0.316; -0.304; -0.601
SN	3	0.900	0.758	0.871	-0.316; 0.305; 0.059; 0.570
PBC	3	0.916	0.786	0.886	-0.316; 0.305; -0.069; 0.415
I	3	0.864	0.691	0.831	-0.601; 0.570; 0.415; 0.133

P = Opinion Polarization; A= Attitude; SN = Subjective Norms; PBC = Perceived Behavioral Control; I = Intention.

Choice of the discussion issue

In the first questionnaire, students were asked to rank issues from the most interesting to the least interesting. The results of this question are reported in fig. 12 (left side).

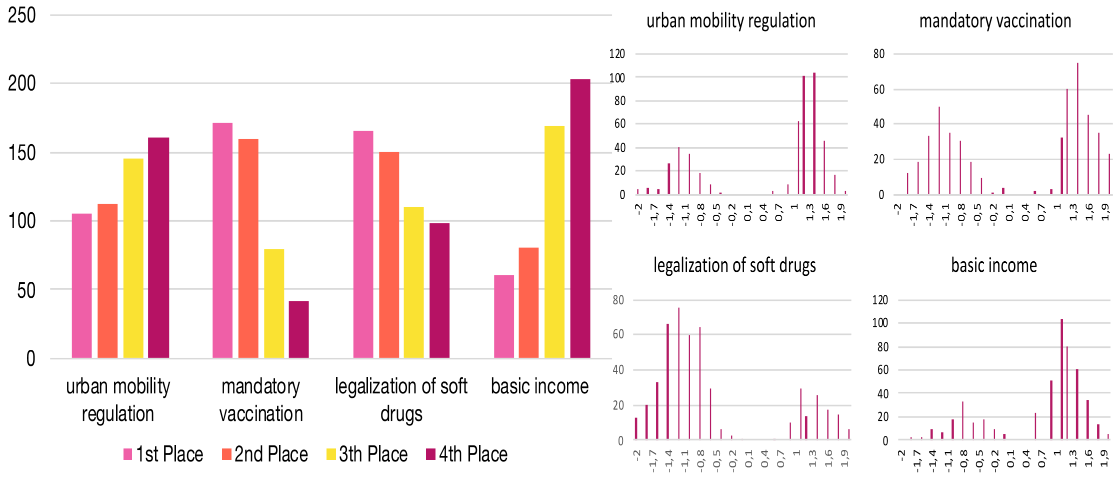


Figure 12: Attractiveness and distribution of opinion about moral dilemmas

34% of students selected mandatory vaccination as the most interesting topic, compared to 33% who selected the legalization of soft drugs. Basic income (12%) and urban mobility regulation (21%) were instead considered less engaging.

The polarizing effect of each topic was evaluated through an index proposed by Morales et al., (2015):

$$\mu = (1 - \Delta A)d$$

where measures the difference between the area of the distribution under each peak and d is the distance between the peaks. With a value equal to 1 Vaccination scored the highest, followed by legalization of soft drugs (0.84), Urban mobility restriction (0.76), and universal income (0.42) Based on these results, Vaccination proved to be both popular and divisive and was therefore selected as the issue for the online discussion step.

Nested-ANOVA and measure of nonindependence

Following Kenny, Kashy, and Bolger (1998), we tested for statistically significant interclass correlation (ρ) for each construct in order to verify the presence of group effects.

A significant value of interclass correlation implies that there is a nonindependence of data within the groups, and therefore groups should be used as unit of analysis in the hypotheses test. The results reported in Tab. 12 show that nonindependence is significantly high for most constructs (see Appendix B for more details). For this reason, the group was used as unit of analysis in the test of the hypotheses.

Table 12: Results of nested-ANOVA for choice the unit of analysis

NULL HYPOTHESIS: THERE IS NO GROUP EFFECT ON INDIVIDUAL JUDGMENTS		
Factor	Levels	Values
Treatment	2	Like minded participants ("Echo-chamber") Cross-minded participants ("Heated debate")
Group (Treatment)	20	Echo: groups from 1 to 10, Debate: groups 11 to 20

$P_{shift} = P_{post} - P_{pre}$	*DF*	*SS*	*MS*	*F-Value*	*P-Value (P-Value adjusted)*
Source	1	4.083	4.083	15.029	0.000*
Treatment					
Group(Treatment)	18	98.044	5.447	20.047	0.000*
Error					
Person(Group(Treatment))	254	69.008	0.272		
Person(Treatment)	272	167.052	0.614	6.649	0.010 (0.02)*
ρ	0.594				
Fratio	20.047*				

$A_{shift} = A_{post} - A_{pre}$	*DF*	*SS*	*MS*	*F-Value*	*P-Value (P-Value adjusted)*
Source	1	35.572	35.572	17.292	0.000*
Treatment					
Group(Treatment)	18	89.242	4.958	2.410	0.001*
Error					
Person(Group(Treatment))	254	522.494	2.057		
Person(Treatment)	272	611.736	2.249	15.817	0.000 (0.008)*
ρ	0.098				
Fratio	2.41*				

$SN_{shift} = SN_{post} - SN_{pre}$	*DF*	*SS*	*MS*	*F-Value*	*P-Value (P-Value adjusted)*
Source	1	0.533	0.533	0.249	0.618
Treatment					
Group(Treatment)	18	158.751	8.820	4.131	0.000*
Error					
Person(Group(Treatment))	254	542.278	2.135		
Person(Treatment)	272	701.029	2.577	0.207	0.650 (0.805)
ρ	0.194				
Fratio	4.13*				

$PBC_{shift} = PBC_{post} - PBC_{pre}$	*DF*	*SS*	*MS*	*F-Value*	*P-Value (P-Value adjusted)*
Source	1	13.627	13.627	7.776	0.006*
Treatment					
Group(Treatment)	18	139.260	7.737	4.415	0.000*
Error					
Person(Group(Treatment))	254	445.145	1.753		
Person(Treatment)	272	584.405	2.149	6.342	0.012 (0.184)
ρ	0.208				
Fratio	4.415*				

$I_{shift} = I_{post} - I_{pre}$	*DF*	*SS*	*MS*	*F-Value*	*P-Value (P-Value adjusted)*
Source	1	0.195	0.195	0.149	0.700
Treatment					
Group(Treatment)	18	211.430	11.746	8.964	0.000*
Error					
Person(Group(Treatment))	254	332.818	1.310		
Person(Treatment)	272	544.248	2.001	0.098	0.755 (0.897)
ρ	0.3799				
Fratio	8.964*				

* significant (for p=0.05)

Online interaction and group polarization

Hp 1.1 and 1.2 predict an increase of opinion polarization consequent to the participation in an online discussion on the controversial issue of mandatory vaccination. Using groups as unit of analysis and Wilcoxon's test for paired samples (Tab.13), we found that there was no polarization in groups composed by like-minded individuals, while opinion polarization did occur for groups in the cross-minded condition (p = 0.05).

Table 13: Results of the Wilcoxon non-parametric test for group polarization

Hp 1.1 Exposure to online Discussions on controversial issue among like-minded participants leads towards opinion radicalization

Hp 1.2 Exposure to online Discussions on controversial issue among cross-minded participants leads towards opinion radicalization

Wilcoxon non-parametric Test

Hypothesis definition

Null hypothesis: $\mu_{Ppost} - \mu_{Ppre} = 0$

Alternative hypothesis: $\mu_{Ppost} - \mu_{Ppre} > 0$

	N	μ_{Ppost}	μ_{Ppre}	*Wilcoxon statistic*	*p-value*
Average individual opinion polarization across groups with like-minded peers	10	0.97	0.98	19.0	0.821
Average individual opinion polarization across groups with cross-minded peers	10	0.60	1.29	55.0	0.003*

* significant (for p=0.05)

We also computed the opinion polarization metric for the individuals belonging to the offline group, i.e., the students who expressed their position on the issue before and after the discussion but did not participate in the online discussion. For this group, no significant shift in opinion was observed during the time window in which the experiment took place. This finding is offered as a control check that can help to support the hypothesis that opinion polarization was endogenous to the discussion.

Impact of participation in online discussion on intention to act

In this section, we report the results for the test of the hypotheses predicting an impact of the participation to biased online discussions on participants' intention to mobilize in support of their cause and on the precursors of such intention as theorized by TPB (Hp 2, 3, 4 and 5).

Table 14 shows that such an effect is not significant at the intention level in both like-minded and cross-minded discussion groups.

Table 14: Results of Wilcoxon non-parametric test for intention

Hp 2.1: Exposure to online Discussions on controversial issue will favor the formation of individual intention to mobilize in groups with like-minded peers.

Hp 2.2: Exposure to online Discussions on controversial issue will favor the formation of individual intention to mobilize in groups with cross-minded peers.

Wilcoxon non-parametric Test

Hypothesis definition

Null hypothesis: $\mu_{Ipost} - \mu_{Ipre} = 0$

Alternative hypothesis: $\mu_{Ipost} - \mu_{Ipre} > 0$

Data	*N*	μ_{Ipost}	μ_{Ipre}	*Wilcoxon statistic*	*P-value*
Average Individual Intention shift across groups with like-minded peers	10	3.23	3.29	26.0	0.581
Average Individual Intention shift across groups with cross-minded peers	10	3.01	3.11	30.0	0.419

* significant (for p=0.05)

The results of tests on the impact of participation in online discussions on intention precursors show a more nuanced picture (Hp 3, 4, and 5 tested in condition 1 = like-minded and 2 = cross-minded).

Significant Attitude shifts were only observed for the groups composed of cross-minded individuals (see tab.15, Hp 3.1. is not confirmed, Hp 3.2 is confirmed).

As for Subjective norm, significant shifts were instead observed in both conditions (tab. 16, Hp 4.1 and 4.2 are both confirmed), while as far as perceived behavioral control, significant shifts were observed only among groups of like-minded people (tab. 17, Hp 5.1. and 5.2 are both confirmed).

Finally, for individuals in the Offline group the data confirm that subjects who did not participate in the online discussion did not change their opinions (0.528), attitude (0.114), subjective norms (0.846), perceived behavioral control (0.840), and intention to act (0.265).

Table 15: Results of Wilcoxon non-parametric test for attitude

Hp 3.1. Exposure to online discussion on a controversial issue increases the degree to which mobilizing in support of the cause is positively valued (attitude) in discussion groups composed on like-minded individuals

Hp 3.2. Exposure to online discussion on a controversial issue increases the degree to which mobilizing in support of the cause is positively valued (attitude) in discussion groups composed on cross-minded individuals

Wilcoxon non-parametric Test

Hypothesis definition

Null hypothesis: $\mu_{Apost} - \mu_{Apre} = 0$

Alternative hypothesis: $\mu_{Apost} - \mu_{Apre} > 0$

	N	μ_{Apost}	μ_{Apre}	*Wilcoxon statistic*	*P-value*
Average Individual Attitude shift across groups with like-minded peers	10	2.01	2.13	36.0	0.207
Average Individual Attitude shift across groups with cross-minded peers	10	2.25	2.99	51.0	0.010*

* significant (for p=0.05)

Table 16: Results of Wilcoxon non-parametric test for subjective norms

Hp 4.1 Exposure to online discussion on a controversial issue increases the degree to which mobilizing in support of the cause is socially acceptable (subjective norms) in groups composed of like-minded individuals

Hp 4.2 Exposure to online discussion on a controversial issue increases the degree to which mobilizing in support of the cause is socially acceptable (subjective norms) in groups composed of cross-minded individuals

Wilcoxon non-parametric Test

Hypothesis definition

Null hypothesis: $\mu_{SNpost} - \mu_{SNpre} = 0$

Alternative hypothesis: $\mu_{SNpost} - \mu_{SNpre} > 0$

	N	μ_{SNpost}	μ_{SNpre}	*Wilcoxon statistic*	*P-value*
Average Individual Subjective Norm shift across groups with like-minded peers	10	3.07	3.60	55.0	0.003*
Average Individual Subjective Norm shift across groups cross-minded peers	10	2.79	3.55	49.0	0.016*

Table 17: Results of Wilcoxon non-parametric test for perceived behavioral control

Hp 5.1 Exposure to online discussion on a controversial issue increase the degree to which mobilizing in support of the cause is perceived as more feasible (perceived behavioral control) in groups composed by like-minded individuals

Hp 5.2 Exposure to online discussion on a controversial issue does not increase the degree to which mobilizing in support of the cause is perceived as more feasible (perceived behavioral control) in groups composed by cross-minded individuals

Wilcoxon non-parametric Test

Hypothesis definition

Null hypothesis: $\mu_{PBCpost} - \mu_{PBCpre} = 0$

Alternative hypothesis: $\mu_{PBCpost} - \mu_{PBCpre} > 0$

	N	$\mu_{PBCpost}$	μ_{PBCpre}	*Wilcoxon statistic*	*P-value*
Average Individual Perceived Behavioral Control shift across groups with like-minded peers	10	3.26	3.46	46.0	0.033*
Average Individual Perceived Behavioral Control shift across groups with cross-minded peers	10	3.22	3.20	42.0	0.077

* significant (for p=0.05)

Discussion

Our study intends to contribute to the ongoing research on whether the production and consumption of information through online conversational interaction has an impact on the polarization of participants' opinions on controversial issues and on the intention to mobilize to support the preferred option. Online discussions, especially after the advent of SM, have been often accused of worsening the quality of online debate by inducing participants to become more radical in their positions. By reviewing existing research on the topic, we have identified two diverse theoretical mechanisms to explain this phenomenon:

- *the creation of information cocoons through interaction among like-minded (radicalizing echo-chamber theory);*

- *the emergence of heated debates in which opposite factions end up fighting against each other (social identity and social argumentation theory).*

Through this study, we performed a systematic empirical comparison between the polarizing predictive power of these two theories.

The first motivation behind this study was to investigate whether the theory of the radicalizing echo-chamber had empirical ground. Barring the limitations of the study that we will discuss later in the conclusions section, our results provides evidence that polarization does occur as a product of online discussions, but not to the extent and in the ways theorized by the proponents of the theory.

Our data show that members of groups of like-minded peers became more extreme in two out of three of the intention determinants: subjective norms and perceived behavioral control. On the contrary, these participants did not develop more extreme attitudes and intention to mobilize. So, opinion polarization was not likely to intensify participants' commitment to supporting the cause actively. If these groups created an echo-chamber at all, they did so by increasing group cohesion through higher conformism and by seeking group's support and encouragement.

Opinion polarization surfaced more strongly among the groups composed of members with opposite positions on the issue. These individuals became more radical in their attitude and subjective norms, but, similarly to the participants in the like-minded groups, participation in an online discussion did not lead to a higher intention to mobilize offline. For these subjects, the group was more of a collective armor that could be brandished against the opposing faction to impose the group's position and preserve its identity.

Thus, social argumentation theory (Mercier and Sperber, 2011) predicts correctly that participants of cross-minded groups would develop more extreme opinions because they engage in confrontation with members bearing the opposite point of view. Social argumentation theory anticipates that opinion polarization takes place by virtue of argumentative conflict with members of the antagonist group. While threats to group identity favor high subjective norms, the focus on defending and attacking leaves little time and energy for group maintenance and problem-solving. That, coupled with potentially diminishing confidence due to exposure to opposite beliefs and attacks, can weaken perceived behavioral control.

On the other hand, participants in like-minded groups do not need to spend energy in rhetorical fights and focus on preserving internal cohesion, encouragement, positive feedback and other types of supportive behavior that can favor a higher level of perceived behavioral control but does not translate in the strengthening of individual opinions and attitudes toward the issue.

Based on our findings, we argue that these two basic group configurations serve two complementary purposes: reinforcement of group cohesion and access to group collective intelligence (echo-chamber) VS leveraging peers' collective strength to support the affirmation of preferred values (battlefield). Since online access makes it easier than ever to have direct access to both situations, it is plausible to expect that such increased double exposure eventually can favor increasing polarization and radicalization of online discourse.

Additionally, real-world participants can be intrinsically motivated and are subject to a much higher level of exposure to disagreeable others, e.g., via the Exposure-News-Discussion framework proposed by Settle (2019). Thus, our controlled study, in which some of these forces were either absent or neutralized while allowing a direct test between the echo-chamber theory, may underestimate the polarization level that can be achieved "in the wild." Of course, additional research and evidence are needed to test this hypothesis.

Conclusion

In this chapter, we focused on the impact of participation in online conversations on controversial topics on opinion polarization and intention formation. Our evidence shows that opinion polarization is more strongly driven by ideological diversity among the participants. Group of like-minded individuals tend to become more radical at the level of some intention determinants (subjective norms, perceived behavioral control) but do not become more extreme in their position on the issue, whereas groups of cross-minded participants tend to radicalize in opinions, attitudes, and subjective norms. In neither condition, the participation in controversial online discussions had any impact on the intention to mobilize. We explained these results through theories of social argumentation.

Limitations

Our study suffers of the typical limitation of a controlled experiment using subjects that were compensated for their participation. While we measured the level of interest and the degree of divisiveness of different topics and picked up issues that scored well on both dimensions, the subjects were not necessarily opinionated and intrinsically motivated. On the other hand, focusing on such users allowed us to observe the occurrence of opinion and intention shifts in a generalist and not already very polarized population that might be more representative of the general public.

The presence of group effects forced us to consider the discussion group as unit of analysis as opposed to individuals, which certainly reduce the degrees of freedom in the study. However, as shown by (Kenny et al., 1998), the loss in power due to the reduction in the degrees of freedom is offset by an increase in the effect size when the interclass correlation is small or moderate (as in our study). Moreover, the price to pay when using groups as units of analysis is definitely smaller than the penalty in terms of inflated significance deriving from using individual observations within groups.

Thus, we encourage other scholars to pursue more rigorous group-based research and to be more cautious about results that are obtained through individual observations.

It is still debated if political polarization is a product of group dynamics that are magnified or facilitated by online interaction or whether what happens online merely reflects an increasing divide in society (Boxell et al., 2017a, 2017b). This is a fascinating topic that would benefit from more systematic comparative studies. However, regardless of whether individuals are already polarized offline, our study offers convincing evidence that participation in online discussions can reinforce polarization in several ways, in particular via exposure in online antagonist points of view.

Implications

Internet is often accused of increasing divisiveness and polarization in political discourse. Our results confirm these expectations by showing that online conversations on controversial topics can favor radicalization to some extent, in particular among participants engaging in conversational fights.

Our findings can be offered to explain why in the current political landscape, more extreme positions and louder voices seem to be more effective at influencing voters. According to the traditional theory of the median voter (Hotelling, 1929), candidates are more likely to win an election if they are able to attract consensus at the center of the political continuum between opposite orientations (e.g., left-right, liberal-conservative). This was definitely true when political communication was delivered primarily through a few generalist mass media, such as main newspapers and a few TV channels. With the Internet, instead, tactics based on the micro-targeting of specific electoral niches based on direct provocation, use of inflammatory language, and even polarizing bots spreading controversial and biased information can resort to the effect of making moderate leaning voters residing in swinging electoral precincts more extreme than they would otherwise be. Some existing studies already provide empirical evidence of this phenomenon (Bail et al., 2018)

Our study then provides a heightened warning about the risk of manipulations of online conversations through the intentional provision of biased and controversial information. These manipulations attempts are clearly on the rise, as shown by the explicit use of polarizing communication on the Internet by some politicians or in ways that are not transparent to users through "partisan" chatbots that automatically post controversial or even fake content (Shao et al., 2017).

Our work provides some suggestions on how to counteract this trend.

First, if cross-minded interaction ends up making participants even more biased towards their initial ideological preference, well-intentioned initiatives aimed at breaking filter bubbles to create more balanced and less biased discussions, perhaps through the injection of scientific and authoritative information, the risk to backfire and produce even more polarized debate (Bessi et al., 2016b; Nyhan and Reifler, 2010; Sunstein et al., 2017).

Second, while we are against any restriction to freedom of speech and affiliation on the Internet, more effort and more effective mechanisms to stop spamming caused by artificial agents impersonating real users should be provided by companies that offer online conversational spaces such as SM companies. Not only are these agents fake, but our results show that when used to spread controversial and fake news, they contribute to increasing intention polarization.

If opinion polarization occurs (even if just at the level of some intention determinants such as attitude), tipping users' opinion even just a little toward either side of controversy can have tremendous consequences in elections in majoritarian political systems or referendums in which even relatively few votes in key districts can reverse the outcome of the election. As the Cambridge Analytical scandal showed, misappropriation of digital data related to users' personality and social network profiles, combined with promotional intervention microtargeting "persuadable users" often exposing them to the "dire" consequences that would follow should the other party prevail, can be a deadly mix to manipulate in totally opaque ways the quality of the information we consume and, thus, the quality of the choices we make. While debate polarization is not the only cause and probably not the most important factor in determining consequences of such magnitude, our study shows that it can certainly have a key role in influencing users' behavior. Therefore, we hope that this paper will motivate other researchers to investigate the fascinating relationship between online information consumption, group dynamics, and impact on actual users' behavior in the real world.

CHAPTER FOUR

GENERAL CONCLUSIONS

In this chapter, we outline the main conclusions of the thesis and we sketch some possible future works in the same direction. This research work investigated the role of SM on the development of GP and peoples' intention to mobilize in sustaining a cause. The first research question that motivated this study was:

RQ1: How do social media contribute to the emergence of group polarization?

Our literature review showed that the popular argument commonly put forth as an explanation of GP SM-enabled is related to their ability to foster the emergence of echo-chambers where radical ideas are amplified. Sunstein (2001), argues that the main characteristic of social networking sites is that they allow like-minded people to find one another. In such an environment, users are only exposed to information that reinforces their own views and remain isolated from challenging one, in part due to the filtering work of algorithms that generate filter bubbles (Pariser, 2011) and human cognitive bias such as the confirmation bias. The outcome of this process is a community that is increasingly segregated along ideological lines, and where compromise becomes unlikely due to rising mistrust on the other side of the ideological spectrum.

Despite an apparent consensus, other studies offer a much more nuanced view of how SM affects GP, often questioning the basic premises of this argument. Even if most exchanges on SM take place among people with similar ideas, cross-cutting interactions are more frequent than commonly believed (Barberá et al., 2015), exposure to diverse news is higher than through other types of media, and ranking algorithms do not have a large impact on online news consumption (Bakshy et al., 2015). A potential explanation for this series of findings is that social networking sites increase exposure to information shared by weak ties, such as co-workers, relatives, and acquaintances, who are more likely to share different points of view (Barberá, 2014).

Of course, the fact that SM increases exposure to diverse ideas from weak ties does not necessarily mean it has no effect on GP. Past research shows that repeated exposure to cross-cutting information leads to moderation (Mutz and Mondak, 2006), which could explain why GP has actually increased the least among those old people who are least likely to use SM (Boxell et al., 2017a). However, a growing body of study challenges this finding, arguing that it is precisely this increased exposure to cross-cutting views that may be having polarizing effects (Bail et al., 2018; Suhay et al., 2018). SM provide opportunities to directly engage one's opponents, to instigate a virtual fight through online confrontation (Merry, 2016). These cross-cutting interactions allow SM users to reinforce in-group and out-group affiliation (Morin and Flynn, 2014; Yardi and Boyd, 2010). Described as a backfire effect, criticism encountered online allows people to develop a greater trust towards parties to which they feel they belong to, and more repudiation opponents (Suhay et al., 2018). Finally, as Zhu and his colleagues point out, selective exposure and homophily can be the consequence of exposure to diversity. Thus, the construction of isolated environments such as those described by the echo-chamber metaphor can be the result of a prior condition of cross exposure that the users themselves refuse or avoid (Zhu et al., 2017). Then, we provide evidence that although the causes and mechanisms proposed to explain the online group dynamic do not greatly different from those that guide the phenomenon in face-to-face interaction, the online environment offers more opportunities to design spaces that are more polarization-conducive. For example, if we consider the echo-chambers hypothesis, GP is explained by causes such as selective exposure and homophily, translating theories used to explore the same phenomenon in the offline context. But, it is easy to understand that offline is not so simple to design spaces with characteristics of scale, reachability, and immediacy similar to online echo-chambers; and the same happens for the cross-cutting condition.

In conclusion, we see that the existing literature offers results that appear to be at odds. Therefore, we wondered:

> *RQ2: Are social media conducive to biased information consumption and suppression of diversity through the emergence of echo-chambers and information cocoons?*

RQ3: Do social media contribute to making group interaction increasingly divisive and fragmented where each point of view is perceived as superior and non-negotiable?

Specifically, we addressed at verifying:

- *Whether online exposure to polarized debate has an impact on people's beliefs and intentions formation*
- *What is the role of online diversity in the development of group polarization and peoples' intention to act.*

The results of our first experiment provide evidence that conversational interaction across members of antagonist groups and, thus, exposure to diverse opinions do occur in online political debates taking place through Twitter backchanneling conversations. In all the three analyzed debates, this conversational interaction evolved in time with a similar trend by increasing during the debate compared to the interaction taking place before and after the debate itself. As predicted by common ground and persuasive argument theories, Twitter users do engage with diverse opinions. However, this interaction seems more dictated by the occurrence of a polarizing event (the debate) than by the need to access new information and possibly reconsider existing beliefs. While Twitter allows users to interact with virtually any other participant, our data show that this openness can backfire, at least when it comes to political discussions: instead of favoring revision of pre-existing beliefs, interaction with other users provoked both antagonist segregation.

While concepts such as filter bubbles or echo chambers have proven problematic to define and identify, we think that tracking conversational interaction across ideologically segregated communities could be a promising perspective in the analysis of this phenomenon. The bubble metaphor can be applied to our results as well in a dynamic way since the fragmentation metric we adopt in this work reveals the inflating and deflating of ideologically homogenous retweet sub-networks.

Therefore, the expectation that online conversations should favor exposure to diverse beliefs is supported by our data; however, the behavior following this exposure does not seem conducive to a more open and fertile debate. Not only the temporary increase is followed by an decrease in interaction across the two groups, but, upon closer inspection of the content of the tweets, we found out that the interaction among users belonging to antagonist factions is typically oriented at attacking opposite points of view or at reasserting prejudicial positions, as predicted by the persuasive argument theory. In conclusion, our results suggest that online ideologically, segregation is the outcome of the combination of two factors: predominance of interaction with like-minded peers coupled with antagonist interaction with the opposite faction during polarizing events.

Finally, through the second study, we performed a systematic empirical comparison between the polarizing predictive power of these two theories. Our data show that members of groups of like-minded peers became more extreme in two out of three of the intention determinants: subjective norms and perceived behavioral control. On the contrary, these participants did not develop more extreme attitudes and intention to mobilize. So, GP was not likely to intensify participants' commitment to supporting the cause actively. If these groups created an echo-chamber at all, they did so by increasing group cohesion through higher conformism and by seeking group's support and encouragement. GP surfaced more strongly among the groups composed of members with opposite positions on the issue. These individuals became more radical in their attitude and subjective norms, but, similarly to the participants in the like-minded groups, they did not lead to a higher intention to mobilize offline. Moreover, our results also showed that threats to group identity favor high subjective norms and that the focus on defending and attacking leaves little time and energy for group maintenance and problem-solving. That, coupled with potentially diminishing confidence due to exposure to opposite beliefs and attacks, can weaken perceived behavioral control. On the other hand, participants in like-minded groups do not need to spend energy in rhetorical fights and focus on preserving internal cohesion, encouragement, positive feedback and other types of supportive behavior that can favor a higher level of perceived behavioral control but does not translate in the strengthening of individual opinions and attitudes toward the issue.

Based on our findings, we argue that these two basic group configurations serve two complementary purposes: reinforcement of group cohesion and access to group collective intelligence (echo-chamber) VS leveraging peers' collective strength to support the affirmation of preferred values (battlefield). Since online access makes it easier than ever to have direct access to both situations, it is plausible to expect that such increased double exposure eventually can favor increasing polarization and radicalization of online discourse.

In conclusion, while our results seem to mitigate the worrying echo-chambers consequences, on the other hand, they demonstrate that it would be naïve to believe that greater diversity could attenuate the emergence of GP or surely improve decision-making. Of course, additional research and evidence are needed to test this hypothesis as described below.

Limitations

Before discussing the implications of our findings, we first note the essential limitations of this research work. Proceeding by order, we are aware that the systematic literature review, presented in the first chapter, has a limitation about its updating date (2018), and the lack of a meta-analysis to compare the results obtained in the 92 selected works.

Subsequently, as the majority of empirical studies, the results of the two experiments (in the second and third chapter) suffer typical limitations that concern the nature of data and the proposed models of analysis. These constraints affect both replication and generalization of our results (that is a typical limitation in these kind of analyses). Regarding the data, in the first experiment, our sample has been limited by the decision to collect Tweets using the free search tool supported by Twitter API. This free API downloads provided us only a random sample of all the messages that American users shared during the investigated events.

Another source of uncertainty is associated to the fact that the collected data do not include information about the network of users' followers, biography, or information related to the total number of likes or retweets that a post/user collected. These limitations are also added to those related to the manual classification that we made to detect elite users. In contrast, in the second experiment, we had more control over the entire data acquisition process (thanks to the detailed experimental design), but our data were affected by the limitations of self-report, typical of survey data.

However, while in both experiments, the generalizability of results is equally limited to the topic of discussion (policy of the two American candidates, and the existence of mandatory vaccination in Italy), regarding the replicability with a larger sample, the second experiment presents higher obstacles. In particular, these obstacles are mainly linked to the sequentiality of the collection phases (pre-discussion questionnaire, group formation, discussion, and post-discussion questionnaire) and the impossibility to exchange or bypass one of these phases without modifying the entire experimental design.

Moreover, regarding conceptual limitations, neither of the two experiments investigated if there was a causal relationship between the observed phenomena and (individual or collective) actions. Indeed, the first experiment does not explore if online GP involves actions of voting for one candidate, while the second experiment did not verify if the occurred shift of intentions and its precursors resulted in a real mobilization.

Finally, our studies are related to the fact that we do not systematically compare online and offline situations; thus, we cannot ascertain whether GP and influence of group interaction are more intense in online or offline settings.

Implications of research findings

Despite these limitations, our results have important implications for policy, theory, and practice. First of all, the policy-makers currently engaged in proposing initiatives aimed at regulating the exploitation of online user data for manipulative purposes can take advantage of the results of this research to propose tools able to monitor and prevent the emergence of GP and its adverse effects. In this case, our results do not support proposed initiatives of breaking the filter bubbles or injection of diversity into the echo chambers. In fact, according to previous studies (Nyhan and Reifler, 2010; Sunstein et al., 2017), our results show that these initiatives could be completely ineffective, if not downright counterproductive. Regulation initiatives should take into account that, although different, both conditions of exposure to online diversity can influence people's decision-making. Thus, policy-makers should prompt the acquisition and development of applications ready to guarantee a more balanced online exposure (between echo-chambers and debate) in order to improve the management of critical situations such as manipulations, subversions, and fanaticisms.

Although it does not often happen to trace positive aspects of GP, the paradox can be significant and useful in providing several implications, for example, in the fields of marketing, healthcare, and welfare in general. Concerning marketing, sometimes, as noted by (Luo et al., 2013), a product or brand is not inherently polarizing, but marketers may want to introduce polarization in order to differentiate it from a strong competitor or to make it stand apart from a crowded field. Thus, brand managers could exploit our approach and our results to better understand how to capitalize online GP and conquer the market. As in politics, the effects of GP (group cohesion, construction of strong ideologies, defense towards others) can be useful to develop loyal and stable brand communities.

Similarly, the influence of SM and polarization on the decision-making process that guides people to adopt certain behaviors, highlighted by our results, could offer valid ideas for better the spread of environmentally friendly and healthy behaviors — exploiting mechanisms similar to those used in micro-targeting our results could be used to design tools aimed to increase people's awareness of particular issues (such as the importance of mandatory vaccinations, drugs abuses, or climate changes) and to persuade them to take actions.

Directions for future research

The present thesis opens up different research directions. In terms of future work, we intend to update the systematic review of the literature by integrating more recent works focused on the relationship between GP and SM. Moreover, we intend to develop a meta-analysis of the results obtained from the revised researches, in order to better appreciate differences in empirical evidences.

In addition, since the methodology used in both empirical experiments, is versatile enough to be applied in several different areas, e.g. marketing and CRM, it would be interesting to repeat the experiments using different GP measurements, discussion topics on controversial issues, and experimental settings (such as the free data collection, the elite classification, or the pre-post design) in order to probe the solidity and limitations of our findings.

Further research is required to fully understand how to exploit information related to the emergence of GP to predict the follow-up that occurs in people's real actions. This will involve the development of investigative models to explore the causal relationship between the emergence of online GP and offline shifts of people actions. We also could consider more advanced text analytics such as argumentation mining, and reflective writing analytics to gain more insight into the way that users are writing (Cabrio and Villata, 2018; Gibson et al., 2017) .

Finally, we argue that while GP and SM exposure are not the only cause and probably not the most important factor in determining consequences of such magnitude, our study shows that they can certainly have a key role in influencing users' behavior, and we hope that this work will motivate other researchers to investigate the fascinating relationship between online information consumption, group dynamics, and impact on people's behavior in the real world.

REFERENCES

Adamic, L.A., Glance, N., 2005. The political blogosphere and the 2004 U.S. election, in: Proceedings of the 3rd International Workshop on Link Discovery - LinkKDD '05. ACM Press, New York, New York, USA, pp. 36–43. https://doi.org/10.1145/1134271.1134277

Ajzen, I., 2004. Constructing a TPB questionnaire: Conceptual and methodological considerations. University of Massachusetts, Amherst.

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Decis. Process. 50, 179–211. https://doi.org/https://doi.org/10.1016/0749-5978(91)90020-T

Alamsyah, A., Adityawarman, F., 2017. Hybrid sentiment and network analysis of social opinion polarization, in: 2017 5th International Conference on Information and Communication Technology, ICoIC7 2017. IEEE, pp. 1–6. https://doi.org/10.1109/ICoICT.2017.8074650

Alvaro, J., Alvarado, J., Guelph, C., 2014. Are you a Lover or a Hater ?: The Impact of the Brand Polarization Marketing Strategy on the Miracle Whip Facebook Brand Community. The University of Guelph.

Aronson, E., Wilson, T.D., Akert, R.M., 2002. Social psychology, 4th ed. ed. Prentice Hall.

Asch, S.E., 1963. Effects of Group Pressure upon the Modification and Distortion of Judgements, in: Guetzkow, H. (Ed.), Groups, Leadership and Men: Research in Human Relations. New York: Russell and Russell, pp. 177–190.

Asker, D., Dinas, E., 2019. Thinking Fast and Furious: Emotional Intensity and Opinion Polarization in Online Media. Public Opin. Q. 83, 487–509. https://doi.org/10.1093/poq/nfz042

Assibong, P.A., Wogu, I.A.P., Sholarin, M.A., Misra, S., Damasevičius, R., Sharma, N., 2020. The Politics of Artificial Intelligence Behaviour and Human Rights Violation Issues in the 2016 US Presidential Elections: An Appraisal, in: Advances in Intelligent Systems and Computing. Springer, pp. 295–309. https://doi.org/10.1007/978-981-13-9364-8_22

Babaei, M., Kulshrestha, J., Chakraborty, A., Benevenuto, F., Gummadi, K.P., Weller, A., 2018. Purple Feed: Identifying High Consensus News Posts on Social Media, in: AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. pp. 10–16. https://doi.org/10.1145/3278721.3278761

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Fallin Hunzaker, M.B., Lee, J., Mann, M., Merhout, F., Volfovsky, A., 2018. Exposure to opposing views on social media can increase political polarization, in: Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, pp. 9216–9221. https://doi.org/10.1073/pnas.1804840115

Bakshy, E., Messing, S., Adamic, L.A., 2015. Exposure to ideologically diverse news and opinion on Facebook. Science (80-.). 348, 1130–1132. https://doi.org/10.1126/science.aaa1160

Balcells, J., Padró-Solanet, A., 2016. Tweeting on Catalonia's independence : The dynamics of political discussion and group polarisation. Medijske Stud. 7, 124–141. https://doi.org/10.20901/ms.7.14.9

Barberá, P., 2014. How Social Media Reduces Mass Political Polarization . Evidence from Germany, Spain, and the U.S., Job Market Paper, New York University.

Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., Bonneau, R., 2015. Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? Psychol. Sci. 26, 1531–1542. https://doi.org/10.1177/0956797615594620

Baum, M.A., Groeling, T., 2008. New Media and the Polarization of American Political Discourse. Polit. Commun. 25, 345–365. https://doi.org/10.1080/10584600802426965

Beam, M.A., Hutchens, M.J., Hmielowski, J.D., 2018. Facebook news and (de)polarization: reinforcing spirals in the 2016 US election. Inf. Commun. Soc. 21, 940–958. https://doi.org/10.1080/1369118X.2018.1444783

Bechterew, W., De Lange, M., 1924. Die Ergebnisse des Experiments auf dem Gebiete der kollektiven Reflexologie. Zsch. f. angew. Psychol 24, 305–344.

Ben-David, A., Matamoros-Fernandez, A., 2016. Hate Speech and Covert Discrimination on Social Media : Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain. Int. J. Commun. 10, 1167–1193. https://doi.org/1932–8036/20160005

Benkler, Y., 2006. The wealth of networks : how social production transforms markets and freedom. Yale University Press.

Bessi, A., 2016. Personality traits and echo chambers on facebook. Comput. Human Behav. 65, 319–324. https://doi.org/10.1016/j.chb.2016.08.016

Bessi, A., Petroni, F., Del Vicario, M., Zollo, F., Anagnostopoulos, A., Scala, A., Caldarelli, G., Quattrociocchi, W., 2016a. Homophily and polarization in the age of misinformation. Eur. Phys. J. Spec. Top. 225, 2047–2059. https://doi.org/10.1140/epjst/e2015-50319-0

Bessi, A., Petroni, F., Del Vicario, M., Zollo, F., Anagnostopoulos, A., Scala, A., Caldarelli, G., Quattrociocchi, W., 2015. Viral misinformation: The role of homophily and polarization, in: Proceedings of the 24th International Conference on World Wide Web. ACM. Association for Computing Machinery, Inc, pp. 355–356. https://doi.org/10.1145/2740908.2745939

Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., Quattrociocchi, W., 2016b. Users polarization on Facebook and Youtube. PLoS One 11. https://doi.org/10.1371/journal.pone.0159641

Bode, L., Hanna, A., Yang, J., Shah, D.V., 2015. Candidate Networks, Citizen Clusters, and Political Expression: Strategic Hashtag Use in the 2010 Midterms. Ann. Am. Acad. Pol. Soc. Sci. 659, 149–165. https://doi.org/10.1177/0002716214563923

Bor, S.E., 2014. Using Social Network Sites to Improve Communication Between Political Campaigns and Citizens in the 2012 Election. Am. Behav. Sci. 58, 1195–1213. https://doi.org/10.1177/0002764213490698

Borge-Holthoefer, J., Magdy, W., Darwish, K., Weber, I., 2015. Content and network dynamics behind Egyptian political polarization on twitter, in: CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing. Association for Computing Machinery, Inc, New York, New York, USA, pp. 700–711. https://doi.org/10.1145/2675133.2675163

Boxell, L., Gentzkow, M., Shapiro, J.M., 2017a. Greater Internet use is not associated with faster growth in political polarization among US demographic groups, in: Proceedings of the National Academy of Sciences. National Academy of Sciences, pp. 10612–10617. https://doi.org/10.1073/pnas.1706588114

Boxell, L., Gentzkow, M., Shapiro, J.M., 2017b. Is the internet causing political polarization? Evidence from demographics.

Bozdag, E., Gao, Q., Houben, G.J., Warnier, M., 2014. Does offline political segregation affect the filter bubble? An empirical analysis of information diversity for Dutch and Turkish Twitter users. Comput. Human Behav. 41, 405–415. https://doi.org/10.1016/j.chb.2014.05.028

Brady, W.J., Wills, J.A., Jost, J.T., Tucker, J.A., Van Bavel, J.J., Fiske, S.T., 2017. Emotion shapes the diffusion of moralized content in social networks, in: Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, Department of Psychology, New York University, New York, NY 10003, United States, pp. 7313–7318. https://doi.org/10.1073/pnas.1618923114

Bravo, R.B., Del Valle, M.E., Gavidia, A.R., 2016. A multilayered analysis of polarization and leaderships in the Catalan Parliamentarians' Twitter Network, in: 15th International Conference on Advances in ICT for Emerging Regions, ICTer 2015 - Conference Proceedings. Institute of Electrical and Electronics Engineers Inc., Open University of Catalonia (UOC), Internet Interdisciplinary Institute (IN3), Parc Mediterrani de la Tecnologia (Edifici B3), Av. Carl Friedrich Gauss, 5, Castelldefels, 08860, Spain, pp. 200–206. https://doi.org/10.1109/ICTER.2015.7377689

Bright, J., 2018. Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. J. Comput. Commun. 23, 17–33. https://doi.org/10.1093/jcmc/zmx002

Bruns, A., 2019. After the 'APIcalypse': social media platforms and their fight against critical scholarly research. Inf. Commun. Soc. 22, 1544–1566. https://doi.org/10.1080/1369118X.2019.1637447

Bruns, A., Highfield, T., Burgess, J., 2013. The Arab Spring and Social Media Audiences: English and Arabic Twitter Users and Their Networks. Am. Behav. Sci. 57, 871–898. https://doi.org/10.1177/0002764213479374

Burtt, H.E., 1920. Sex Differences in the Effect of Discussion. J. Exp. Psychol. 3, 390–395. https://doi.org/10.1037/h0072937

Cabrio, E., Villata, S., 2018. Five years of argument mining: A Data-driven Analysis, in: IJCAI International Joint Conference on Artificial Intelligence. International Joint

Conferences on Artificial Intelligence Organization, California, pp. 5427–5433. https://doi.org/10.24963/ijcai.2018/766

Castells, M., Fernández-Ardèvol, M., Qiu, L.J., Sey, A., 2009. Mobile Communication and Society: A Global Perspective. The MIT Press., Cambridge, MA.

Ceron, A., Curini, L., Iacus, S.M., Porro, G., 2014. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. New Media Soc. 16, 340–358. https://doi.org/10.1177/1461444813480466

Chan, C.H., Fu, K.W., 2017. The Relationship Between Cyberbalkanization and Opinion Polarization: Time-Series Analysis on Facebook Pages and Opinion Polls During the Hong Kong Occupy Movement and the Associated Debate on Political Reform. J. Comput. Commun. 22, 266–283. https://doi.org/10.1111/jcc4.12192

Chan, C.H., Fu, K.W., 2015. Predicting political polarization from cyberbalkanization: Time series analysis of facebook pages and opinion poll during the Hong Kong occupy movement, in: Proceedings of the 2015 ACM Web Science Conference. https://doi.org/10.1145/2786451.2786509

Cho, J., Ahmed, S., Keum, H., Choi, Y.J., Lee, J.H., 2016. Influencing Myself: Self-Reinforcement Through Online Political Expression. Communic. Res. 45, 83–111. https://doi.org/10.1177/0093650216644020

Clark, H.H., Brennan, S.E., 1991. Grounding in communication., in: Perspectives on Socially Shared Cognition. pp. 127–149. https://doi.org/10.1037/10096-006

Clark, H.H., Schaefer, E.F., 1989. Contributing to discourse. Cogn. Sci. 13, 259–294. https://doi.org/10.1016/0364-0213(89)90008-6

Clauset, A., Eagle, N., 2012. Persistence and periodicity in a dynamic proximity network, in: ArXiv Preprint ArXiv:1211.7343.

Conover, M.D., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., Flammini, A., 2011a. Political polarization on Twitter., in: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media Political. pp. 89–96. https://doi.org/10.1021/ja202932e

Conover, M.D., Ratkiewicz, M.F., Gonçalves, B., Flammini, A., Menczer, F., 2011b. Political Polarization on Twitter, in: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media Political. pp. 10262–10274.

https://doi.org/10.1021/ja202932e

Costa e Silva, E., 2014. A deliberative public sphere? Picturing Portuguese political blogs. Observatorio 8, 187–204.

Dandekar, P., Goel, A., Lee, D.T., 2013. Biased assimilation, homophily, and the dynamics of polarization. Proc. Natl. Acad. Sci. U. S. A. 110, 5791–5796. https://doi.org/10.1073/pnas.1217220110

De Kerckhove, D., 1997. Connected intelligence : the arrival of the Web society. Somerville House, Toronto.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, E.H., Quattrociocchi, W., 2016a. The spreading of misinformation online, in: Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, pp. 554–559. https://doi.org/10.1073/pnas.1517441113

Del Vicario, M., Gaito, S., Quattrociocchi, W., Zignani, M., Zollo, F., 2017a. News consumption during the Italian referendum: A cross-platform analysis on facebook and twitter, in: Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017. pp. 648–657. https://doi.org/10.1109/DSAA.2017.33

Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H.E., Quattrociocchi, W., 2017b. Modeling confirmation bias and polarization. Sci. Rep. 7, 40391. https://doi.org/10.1038/srep40391

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., Quattrociocchi, W., 2016b. Echo Chambers: Emotional Contagion and Group Polarization on Facebook. Sci. Rep. 6, 1–12. https://doi.org/10.1038/srep37825

Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., Quattrociocchi, W., 2017c. Mapping social dynamics on Facebook: The Brexit debate. Soc. Networks 50, 6–16. https://doi.org/10.1016/j.socnet.2017.02.002

Della Porta, D., Mosca, L., 2005. Global-net for global movements? A network of networks for a movement of movements. J. Public Policy 25, 165–190. https://doi.org/10.1017/S0143814X05000255

Dilliplane, S., 2011. All the news you want to hear: The impact of partisan news exposure on political participation. Public Opin. Q. 75, 287–316. https://doi.org/10.1093/poq/nfr006

DiMaggio, P., Evans, J.H., Bryson, B., 1996. Have Americans' attitudes become more polarized? Am. J. Sociol. 102, 690–755. https://doi.org/https://doi.org/10.1086/230995

Du, S., Gregory, S., 2016. The echo chamber effect in twitter: Does community polarization increase?, in: International Workshop on Complex Networks and Their Applications. pp. 373–378. https://doi.org/10.1007/978-3-319-50901-3_30

Dubois, E., Blank, G., 2018. The echo chamber is overstated: the moderating effect of political interest and diverse media. Inf. Commun. Soc. 21, 729–745. https://doi.org/10.1080/1369118X.2018.1428656

Easterby-Smith, M., Thorpe, R., Jackson, P.R., Jaspersen, L.., 2015. Management and business research, 5th Ed. ed.

Elmedni, B., 2016. Death of Rationality: The Social Networks' Factor in Policy Response to Ebola. Int. J. Public Adm. 39, 917–926. https://doi.org/10.1080/01900692.2015.1057851

Enli, G., 2017. Twitter as arena for the authentic outsider: exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election. Eur. J. Commun. 32, 50–61. https://doi.org/10.1177/0267323116682802

Everton, S.F., 2016. Social Networks and Religious Violence. Rev. Relig. Res. 58, 191–217. https://doi.org/10.1007/s13644-015-0240-3

Festinger, L., 1954. A Theory of Social Comparison Processes. Hum. Relations 7, 117–140. https://doi.org/10.1177/001872675400700202

Finn, S., Mustafaraj, E., Metaxas, P.T., 2014. The Co-retweeted Network and Its Applications for Measuring the Perceived Political Polarization, in: Proceedings of the 10th International Conference on Web Information Systems and Technologies. SCITEPRESS - Science and and Technology Publications, Department of Computer Science, Wellesley College, Wellesley, MA 02481, United States, pp. 276–284. https://doi.org/10.5220/0004788702760284

Flaxman, S., Goel, S., Rao, J.M., 2016. Filter bubbles, echo chambers, and online news consumption. Public Opin. Q. 80, 298–320. https://doi.org/10.1093/poq/nfw006

Fornell, C., Larcker, D.F., 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. J. Mark. Res. 18, 39–50. https://doi.org/10.1177/002224378101800104

Francis, J.J., Eccles, M.P., Johnston, M., Walker, A., Grimshaw, J., Foy, R., Kaner, E.F., Smith, L., Bonetti, D., Francis, J., Eccles, M., Kaner, E., 2004. Constructing questionnaires based on the theory of planned behaviour: A manual for health services researchers.

Garcia, D., Abisheva, A., Schweighofer, S., Serdült, U., Schweitzer, F., 2015. Ideological and temporal components of network polarization in online political participatory media. Policy and Internet 7, 46–79. https://doi.org/10.1002/poi3.82

Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M., 2017a. The Ebb and flow of controversial debates on social media, in: Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017. AAAI Press, Aalto University, Helsinki, Finland, pp. 524–527.

Garimella, K., Gionis, A., Morales, G.D.F., Mathioudakis, M., De Francisci Morales, G., Mathioudakis, M., 2017b. The effect of collective atention on controversial debates on social media, in: WebSci 2017 - Proceedings of the 2017 ACM Web Science Conference. pp. 43–52. https://doi.org/10.1145/3091478.3091486

Garimella, K., Weber, I., 2017. A long-term analysis of polarization on Twitter, in: Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017. pp. 528–531.

Garimella, K., Weber, I., De Choudhury, M., 2016. Quote RTs on Twitter: Usage of the new feature for political discourse, in: WebSci 2016 - Proceedings of the 2016 ACM Web Science Conference. Association for Computing Machinery, Inc, Aalto University, Helsinki, Finland, pp. 200–204. https://doi.org/10.1145/2908131.2908170

Garrett, R.K., 2009. Echo chambers online?: Politically motivated selective exposure among Internet news users. J. Comput. Commun. 14, 265–285. https://doi.org/10.1111/j.1083-6101.2009.01440.x

Garrett, R.K., Gvirsman, S.D., Johnson, B.K., Tsfati, Y., Neo, R., Dal, A., 2014. Implications of Pro- and Counterattitudinal Information Exposure for Affective Polarization. Hum. Commun. Res. 40, 309–332. https://doi.org/10.1111/hcre.12028

Gibson, A., Shum, S.B., Aitken, A., Tsingos-Lucas, C., Sándor, Á., Knight, S., 2017. Reflective writing analytics for actionable feedback, in: ACM International Conference Proceeding Series. Association for Computing Machinery, New York, New York, USA, pp. 153–162. https://doi.org/10.1145/3027385.3027436

Gigone, D., Hastie, R., 1993. The Common Knowledge Effect: Information Sharing and Group Judgment. J. Pers. Soc. Psychol. 65, 959–974. https://doi.org/10.1037/0022-3514.65.5.959

Gonzalez-Bailon, S., Borge-Holthoefer, J., Moreno, Y., 2012. Broadcasters and Hidden Influentials in Online Protest Diffusion. SSRN Electron. J. https://doi.org/10.2139/ssrn.2017808

Grabowicz, P.A., Ramasco, J.J., Moro, E., Pujol, J.M., Eguiluz, V.M., 2012. Social features of online networks: The strength of intermediary ties in online social media. PLoS One 7. https://doi.org/10.1371/journal.pone.0029358

Grömping, M., 2014. 'Echo Chambers': Partisan Facebook Groups during the 2014 Thai Election. Asia Pacific Media Educ. 24, 39–59. https://doi.org/10.1177/1326365X14539185

Groshek, J., Koc-Michalska, K., 2017. Helping populism win? Social media use, filter bubbles, and support for populist presidential candidates in the 2016 US election campaign. Inf. Commun. Soc. 20, 1389–1407. https://doi.org/10.1080/1369118X.2017.1329334

Gruzd, A., Roy, J., 2014. Investigating Political Polarization on Twitter: A Canadian Perspective. Policy & Internet 6, 28–45. https://doi.org/10.1002/1944-2866.POI354

Guerra, C.P.., Souza, R.C.S.N.P., Assunção, R.M., Meira, W., 2017. Antagonism also flows through retweets: The impact of out-of-context quotes in opinion polarization analysis, in: Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017. pp. 536–539.

Guerra, C.P.H., Meira, W., Cardie, C., Kleinberg, R., 2013. A measure of polarization on social media NetworksBased on community boundaries, in: Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013. pp. 215–224.

Habibi, M.R., Laroche, M., Richard, M.-O., 2014. Brand communities based in social media: How unique are they? Evidence from two exemplary brand communities. Int. J. Inf. Manage. 34, 123–132. https://doi.org/10.1016/j.ijinfomgt.2013.11.010

Hampton, K., Rainie, L., Lu, W., Dwyer, M., Shin, I., Purcell, K., 2014. Social Media and the 'Spiral of Silence.' Pew Res. Cent.

Hanna, A., Wells, C., Maurer, P., Friedland, L.A., Shah, D., Matthes, J., 2013. Partisan alignments and political polarization online: A computational approach to understanding the French and US presidential elections, in: PLEAD 2013 - Proceedings of the Workshop on Politics, Elections and Data, Co-Located with CIKM 2013. Department of Sociology, University of Wisconsin-Madison, Madison, WI, United States, pp. 15–21. https://doi.org/10.1145/2508436.2508438

Hargittai, E., Gallo, J., Kane, M., 2008. Cross-ideological discussions among conservative and liberal bloggers. Public Choice 134, 67–86. https://doi.org/10.1007/s11127-007-9201-x

He, W., Zha, S., Li, L., 2013. Social media competitive analysis and text mining: A case study in the pizza industry. Int. J. Inf. Manage. 33, 464–472.

Heatherly, K.A., Lu, Y., Lee, J.K., 2017. Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites. New Media Soc. 19, 1271–1289. https://doi.org/10.1177/1461444816634677

Hemphill, L., Culotta, A., Heston, M., 2016. #Polar Scores: Measuring partisanship using social media content. J. Inf. Technol. Polit. 13, 365–377. https://doi.org/10.1080/19331681.2016.1214093

Hong, S., Kim, S.H., 2016. Political polarization on twitter: Implications for the use of social media in digital governments. Gov. Inf. Q. 33, 777–782. https://doi.org/10.1016/j.giq.2016.04.007

Hotelling, H., 1929. Stability in Competition. Econ. J. 39, 41. https://doi.org/10.2307/2224214

Iandoli, L., Quinto, I., De Liddo, A., Buckingham Shum, S., 2014. Socially augmented argumentation tools: Rationale, design and evaluation of a debate dashboard. Int. J. Hum. Comput. Stud. 72, 298–319. https://doi.org/10.1016/j.ijhcs.2013.08.006

Ince, J., Rojas, F., Davis, C.A., 2017. The social media response to Black Lives Matter: how Twitter users interact with Black Lives Matter through hashtag use. Ethn. Racial Stud. 40, 1814–1830. https://doi.org/10.1080/01419870.2017.1334931

Introne, J., Gokce Yildirim, I., Iandoli, L., DeCook, J., Elzeini, S., 2018. How People Weave Online Information Into Pseudoknowledge. Soc. Media Soc. 4. https://doi.org/10.1177/2056305118785639

Isenberg, D.J., 1986. Group Polarization: A Critical Review and Meta-Analysis. J. Pers.

Soc. Psychol. 50, 1141.

Iyengar, S., Hahn, K.S., 2009. Red media, blue media: Evidence of ideological selectivity in media use. J. Commun. 59, 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., Westwood, S.J., 2019. The Origins and Consequences of Affective Polarization in the United States. Annu. Rev. Polit. Sci. 22, 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Johnson, T.J., Kaye, B.K., Lee, A.M., 2017. Blinded by the Spite? Path Model of Political Attitudes, Selectivity, and Social Media. Atl. J. Commun. 25, 181–196. https://doi.org/10.1080/15456870.2017.1324454

Kawasaki, G., 2004. The art of the start : the time-tested, battle-hardened guide for anyone starting anything. Penguin.

Kenny, D.A., Judd, C.M., 1986. Consequences of violating the independence assumption in analysis of variance. Psychol. Bull. 99, 422.

Kenny, D.A., Kashy, D.A., Bolger, N., 1998. Data analysis in social psychology., in: The Handbook of Social Psychology. John Wiley & Sons, pp. 233–265. https://doi.org/10.1002/9780470561119

Kenny, D.A., Mannetti, L., Pierro, A., Livi, S., 2002. The statistical analysis of data from small groups. J. Pers. Soc. Psychol. 83, 126. https://doi.org/10.1037/0022-3514.83.1.126

Khondker, H.H., 2011. Role of the New Media in the Arab Spring. Globalizations 8, 675–679. https://doi.org/10.1080/14747731.2011.621287

Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S., 2011. Social media? Get serious! Understanding the functional building blocks of social media. Bus. Horiz. 54, 241–251. https://doi.org/10.1016/j.bushor.2011.01.005

Kuo, R., 2018. Racial justice activist hashtags: Counterpublics and discourse circulation. New Media Soc. 20, 495–514. https://doi.org/10.1177/1461444816663485

Lai, M., Bosco, C., Patti, V., Virone, D., 2015. Debate on political reforms in Twitter: A hashtag-driven analysis of political polarization, in: Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015. https://doi.org/10.1109/DSAA.2015.7344884

Lamm, H., 1988. A Review of Our Research on Group Polarization: Eleven Experiments on the Effects of Group Discussion on Risk Acceptance, Probability Estimation, and Negotiation Positions. Psychol. Rep. 62, 807–813. https://doi.org/10.2466/pr0.1988.62.3.807

Larose, D.T., 2005. Discovering Knowledge in Data: an introduction to data mining. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/0471687545

Lawrence, E., Sides, J., Farrell, H., 2010. Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics. Perspect. Polit. 8, 141–157. https://doi.org/10.1017/S1537592709992714

Lee, C., Shin, J., Hong, A., 2018. Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in South Korea. Telemat. Informatics 35, 245–254. https://doi.org/10.1016/j.tele.2017.11.005

Lee, F., Chan, J., 2016. Digital media activities and mode of participation in a protest campaign: a study of the Umbrella Movement. Inf. Commun. Soc. 19, 4–22. https://doi.org/10.1080/1369118X.2015.1093530

Lee, F., Chen, H., Chan, M., 2017. Social media use and university students' participation in a large-scale protest campaign: The case of Hong Kong's Umbrella Movement. Telemat. Informatics 34, 457–469. https://doi.org/10.1016/j.tele.2016.08.005

Lee, J., Choi, J., Kim, C., Kim, Y., 2014. Social media, network heterogeneity, and opinion polarization. J. Commun. 64, 702–722. https://doi.org/10.1111/jcom.12077

Levendusky, M.S., 2013. Why do partisan media polarize viewers? Am. J. Pol. Sci. 57, 611–623. https://doi.org/10.1111/ajps.12008

Levy, P., 1998. Collective intelligence: mankind's emerging world in cyberspace. Choice Rev. Online 35, 35-3911-35–3911. https://doi.org/10.5860/choice.35-3911

Lipizzi, C., Iandoli, L., Marquez, J.E.R., 2016. Combining structure, content and meaning in online social networks: The analysis of public's early reaction in social media to newly launched movies. Technol. Forecast. Soc. Change 109, 35–49. https://doi.org/10.1016/J.TECHFORE.2016.05.013

Livingstone, S., Bober, M., Helsper, E.J., 2005. Active participation or just more information? Information, Commun. Soc. 8, 287–314. https://doi.org/10.1080/13691180500259103

Luo, X., Wiles, M., Raithel, S., 2013. Make the Most of a Polarizing Brand. Harv. Bus.

Rev.

Lynch, M., Freelon, D., Aday, S., 2017. Online clustering, fear and uncertainty in Egypt's transition. Democratization 24, 1159–1177. https://doi.org/10.1080/13510347.2017.1289179

Malone, T.W., Klein, M., 2007. Harnessing Collective Intelligence to Address Global Climate Change. Innov. Technol. Governance, Glob. 2, 15–26. https://doi.org/10.1162/itgg.2007.2.3.15

Marozzo, F., Bessi, A., 2018. Analyzing polarization of social media users and news sites during political campaigns. Soc. Netw. Anal. Min. 8. https://doi.org/10.1007/s13278-017-0479-5

Matakos, A., Terzi, E., Tsaparas, P., 2017. Measuring and moderating opinion polarization in social networks. Data Min. Knowl. Discov. 31, 1480–1505. https://doi.org/10.1007/s10618-017-0527-9

Medaglia, R., Yang, Y., 2017. Online public deliberation in China: evolution of interaction patterns and network homophily in the Tianya discussion forum. Inf. Commun. Soc. 20, 733–753. https://doi.org/10.1080/1369118X.2016.1203974

Medaglia, R., Zhu, D., 2017. Public deliberation on government-managed social media: A study on Weibo users in China. Gov. Inf. Q. 34, 533–544. https://doi.org/10.1016/j.giq.2017.05.003

Medaglia, R., Zhu, D., 2016. Paradoxes of deliberative interactions on government-managed social media: Evidence from China, in: Y., K., S.M., L. (Eds.), ACM International Conference Proceeding Series. Association for Computing Machinery, Copenhagen Business School, Howitzvej 60, Frederiksberg, DK-2000, Denmark, pp. 435–444. https://doi.org/10.1145/2912160.2912184

Mercier, H., Sperber, D., 2011. Why do humans reason? Arguments for an argumentative theory. Behav. Brain Sci. 34, 57–74. https://doi.org/10.1017/S0140525X10000968

Merry, M., 2016. Making friends and enemies on social media: The case of gun policy organizations. Online Inf. Rev. 40, 624–642. https://doi.org/10.1108/OIR-10-2015-0333

Messing, S., Westwood, S.J., 2014. Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. Communic. Res. 41, 1042–1063. https://doi.org/10.1177/0093650212466406

Min, H., Yun, S., 2018. Selective exposure and political polarization of public opinion on the presidential impeachment in South Korea: Facebook vs. kakaotalk. Korea Obs. 49, 137–159. https://doi.org/10.29152/KOIKS.2018.49.1.137

Morales, A.J., Borondo, J., Losada, J.C., Benito, R.M., 2015. Measuring political polarization: Twitter shows the two sides of Venezuela. Chaos 25, 033114. https://doi.org/10.1063/1.4913758

Morin, D.T., Flynn, M.A., 2014. We Are the Tea Party!: The Use of Facebook as an Online Political Forum for the Construction and Maintenance of in-Group Identification during the "GOTV" Weekend. Commun. Q. 62, 115–133. https://doi.org/10.1080/01463373.2013.861500

Morran, D.K., Robison, F.F., Hulse-Killacky, D., 1990. Group research and the unit of analysis problem: The use of anova designs with nested factors. J. Spec. Gr. Work 15, 10–14. https://doi.org/10.1080/01933929008411906

Moscovici, S., Doise, W., Dulong, R., 1972. Studies in group decision II: Differences of positions, differences of opinion and group polarization. Eur. J. Soc. Psychol. 2, 385–399. https://doi.org/10.1002/ejsp.2420020404

Mutz, D.C., Mondak, J.J., 2006. The Workplace as a Context for Cross-Cutting Political Discourse. J. Polit. 68, 140–155. https://doi.org/10.1111/j.1468-2508.2006.00376.x

Myers, D.G., Lamm, H., 1976. The group polarization phenomenon. Psychol. Bull. 83, 602–627. https://doi.org/10.1037/0033-2909.83.4.602

Noelle-Neumann, E., 1974. The Spiral of Silence A Theory of Public Opinion. J. Commun. 24, 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Nyhan, B., Reifler, J., 2010. When corrections fail: The persistence of political misperceptions. Polit. Behav. 32, 303–330. https://doi.org/10.1007/s11109-010-9112-2

Pariser, E., 2012. The filter bubble : how the new personalized web is changing what we read and how we think. Penguin UK.

Pariser, E., 2011. The Filter Bubble: What The Internet Is Hiding From You. Penguin.

Park, Y.J., Jang, S.M., Lee, H., Yang, G.S., 2018. Divide in Ferguson: Social Media, Social Context, and Division. Soc. Media Soc. 4. https://doi.org/10.1177/2056305118789630

Parsell, M., 2008. Pernicious virtual communities: Identity, polarisation and the Web 2.0.

Ethics Inf. Technol. 10, 41–56. https://doi.org/10.1007/s10676-008-9153-y

Parsons, B.M., 2010. Social networks and the affective impact of political disagreement. Polit. Behav. 32, 181–204. https://doi.org/10.1007/s11109-009-9100-6

Postmes, T., Spears, R., Lea, M., 2000. The formation of group norms in computer-mediated communication. Hum. Commun. Res. 26, 341–371. https://doi.org/10.1111/j.1468-2958.2000.tb00761.x

Primario, S., Borrelli, D., Zollo, G., Iandoli, L., Lipizzi, C., 2017. Measuring polarization in Twitter enabled in online political conversation: The case of 2016 US Presidential election, in: Proceedings - 2017 IEEE International Conference on Information Reuse and Integration, IRI 2017. pp. 607–613. https://doi.org/10.1109/IRI.2017.73

Prior, M., 2013. Media and Political Polarization. Annu. Rev. Polit. Sci. 16, 101–127. https://doi.org/10.1146/annurev-polisci-100711-135242

Rainie, H., Wellman, B., 2012. Networked : the new social operating system. MIT Press.

Romenskyy, M., Spaiser, V., Ihle, T., Lobaskin, V., 2018. Polarized Ukraine 2014: Opinion and territorial split demonstrated with the bounded confidence XY model, parametrized by Twitter data. R. Soc. Open Sci. 5. https://doi.org/10.1098/rsos.171935

Schein, E., 2010. Organizational Culture and Leadership Defined, 4th Ed. ed.

Semaan, B.C., Robertson, S.P., Douglas, S., Maruyama, M., 2014. Social media supporting political deliberation across multiple public spheres, in: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW ’14. pp. 1409–1421. https://doi.org/10.1145/2531602.2531605

Settle, J.E., 2019. Frenemies: How social media polarizes America, Cambridge University Press.

Shao, C., Ciampaglia, G.L., Varol, O., Flammini, A., Menczer, F., 2017. The spread of fake news by social bots.

Shapiro, R.Y., 2013. Hearing the Opposition: It Starts at the Top. Crit. Rev. 25, 226–244. https://doi.org/10.1080/08913811.2013.843876

Shearer, C., Watson, H.J., Grecich, D.G., Moss, L., Adelman, S., Hammer, K., Herdlein, S. a, 2000. The CRISP-DM model: The New Blueprint for Data Mining. J. Data Warehous. 5, 13–22.

Shin, J., Jian, L., Driscoll, K., Bar, F., 2017. Political rumoring on Twitter during the

2012 US presidential election: Rumor diffusion and correction. New Media Soc. 19, 1214–1235. https://doi.org/10.1177/1461444816634054

Shore, J., Baek, J., Dellarocas, C., 2018. Network structure and patterns of information diversity on twitter. MIS Q. Manag. Inf. Syst. 42, 849–872. https://doi.org/10.25300/MISQ/2018/14558

Sobkowicz, P., Sobkowicz, A., 2012. Two-Year Study of Emotion and Communication Patterns in a Highly Polarized Political Discussion Forum. Soc. Sci. Comput. Rev. 30, 448–469. https://doi.org/10.1177/0894439312436512

Stoner, J.A.F., 1968. Risky and cautious shifts in group decisions: The influence of widely held values. J. Exp. Soc. Psychol. 4, 442–459. https://doi.org/10.1016/0022-1031(68)90069-3

Stoner, J.A.F., 1961. A comparison of individual and group decisions involving risk. Massachusetts Institute of Technology. https://doi.org/http://hdl.handle.net/1721.1/11330

Stroud, N.J., 2010. Polarization and partisan selective exposure. J. Commun. 60, 556–576. https://doi.org/10.1111/j.1460-2466.2010.01497.x

Suhay, E., Bello-Pardo, E., Maurer, B., 2018. The Polarizing Effects of Online Partisan Criticism: Evidence from Two Experiments. Int. J. Press. 23, 95–115. https://doi.org/10.1177/1940161217740697

Sunstein, C.R., 2008. Neither hayek nor habermas. Public Choice 134, 87–95. https://doi.org/10.1007/s11127-007-9202-9

Sunstein, C.R., 2007. The Polarization of Extremes. Chron. High. Educ. 54, B9.

Sunstein, C.R., 2002a. Why they hate us: The role of social dynamics. Harvard J. Law Public Policy 25, 429–440.

Sunstein, C.R., 2002b. The Law of Group Polarization. J. Polit. Philos. 10, 175–195. https://doi.org/10.1002/9780470690734.ch4

Sunstein, C.R., 2001. Republic.com. Princeton University Press.

Sunstein, C.R., Bobadilla-Suarez, S., Lazzaro, S.C., Sharot, T., 2017. How People Update Beliefs about Climate Change: Good News and Bad News. Cornell L. Rev. 102, 1431–1443.

Tajfel, H., 1982. Social Psychology of Intergroup Relations. Annu. Rev. Psychol. 33, 1–39. https://doi.org/https://doi.org/10.1146/annurev.ps.33.020182.000245

Tajfel, H., Turner, J.C., 1979. An integrative theory of intergroup conflict.

Thaler, R.H., Sunstein, C.R., 2009. Nudge : improving decisions about health, wealth, and happiness. Penguin.

Thelwall, M., Haustein, S., Larivière, V., Sugimoto, C.R., 2013. Do Altmetrics Work? Twitter and Ten Other Social Web Services. PLoS One 8. https://doi.org/10.1371/journal.pone.0064841

Thorndike, R.L., 1938. The Effect of Discussion upon the Correctness of Group Decisions, when the Factor of Majority Influence is Allowed For. J. Soc. Psychol. 9, 343–362. https://doi.org/10.1080/00224545.1938.9920036

Tomlinson, J., 2007. The Culture of Speed : The Coming of Immediacy Delivery Chapter 6 : Delivery. SAGE Publications. https://doi.org/10.4135/9781446212738

Törnberg, P., 2018. Echo chambers and viral misinformation: Modeling fake news as complex contagion. PLoS One 13, 1–21. https://doi.org/10.1371/journal.pone.0203958

Tremayne, M., 2014. Anatomy of Protest in the Digital Era: A Network Analysis of Twitter and Occupy Wall Street. Soc. Mov. Stud. 13, 110–126. https://doi.org/10.1080/14742837.2013.830969

Treré, E., Jeppesen, S., Mattoni, A., 2017. Comparing digital protest media imaginaries: Anti-austerity movements in Spain, Italy & Greece. TripleC 15, 404–422. https://doi.org/10.31269/TRIPLEC.V15I2.772

Tucker, J., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., Nyhan, B., 2018. Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. SSRN Electron. J. https://doi.org/10.2139/ssrn.3144139

Turetsky, K.M., Riddle, T.A.A., 2018. Porous Chambers, Echoes of Valence and Stereotypes: A Network Analysis of Online News Coverage Interconnectedness Following a Nationally Polarizing Race-Related Event. Soc. Psychol. Personal. Sci. 9, 163–175. https://doi.org/10.1177/1948550617733519

Turner, J.C., Reynolds, K.J., 2011. Self-Categorization Theory., in: Handbook of Theories of Social Psychology. SAGE Publications Inc., pp. 399–417. https://doi.org/10.4135/9781446249222.n46

Twenge, J.M., Honeycutt, N., Prislin, R., Sherman, R.A., 2016. More Polarized but More

Independent: Political Party Identification and Ideological Self-Categorization Among U.S. Adults, College Students, and Late Adolescents, 1970-2015. Personal. Soc. Psychol. Bull. 42, 1364–1383. https://doi.org/10.1177/0146167216660058

Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J.T., Nagler, J., Tucker, J.A., 2015. Political expression and action on social media: Exploring the relationship between lower- and higher-threshold political activities among twitter users in Italy. J. Comput. Commun. 20, 221–239. https://doi.org/10.1111/jcc4.12108

Van Alstyne, M., Brynjolfsson, E., 2005. Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities. Manage. Sci. 51, 851–868. https://doi.org/10.1287/mnsc.1050.0363

Wallace, P.M., 2001. The psychology of the Internet, 1st Ed. ed.

Wang, Q., Yang, X., Xi, W., 2018. Effects of group arguments on rumor belief and transmission in online communities: An information cascade and group polarization perspective. Inf. Manag. 55, 441–449. https://doi.org/10.1016/j.im.2017.10.004

Warner, B.R., 2010. Segmenting the electorate: The effects of exposure to political extremism online. Commun. Stud. 61, 430–444. https://doi.org/10.1080/10510974.2010.497069

Weber, I., Garimella, V.R.K., Batayneh, A., 2013. Secular vs. Islamist polarization in Egypt on Twitter, in: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 290–297. https://doi.org/10.1145/2492517.2492557

Weinberger, D., 2004. Is there an echo in here? salon.com.

Wells, C., Cramer, K.J., Wagner, M.W., Alvarez, G., Friedland, L.A., Shah, D.V., Bode, L., Edgerly, S., Gabay, I., Franklin, C., 2017. When We Stop Talking Politics: The Maintenance and Closing of Conversation in Contentious Times. J. Commun. 67, 131–157. https://doi.org/10.1111/jcom.12280

Wiest, J.B., 2011. The Arab Spring| Social Media in the Egyptian Revolution: Reconsidering Resource Mobilization Theory. Int. J. Commun. 5, 18.

Williams, H.T.P., McMurray, J.R., Kurz, T., Hugo Lambert, F., 2015. Network analysis reveals open forums and echo chambers in social media discussions of climate change. Glob. Environ. Chang. 32, 126–138. https://doi.org/10.1016/j.gloenvcha.2015.03.006

Wojcieszak, M.E., 2010. “Don’t talk to me”: Effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism. New Media Soc. 12, 637–655. https://doi.org/10.1177/1461444809342775

Wolfsfeld, G., Segev, E., Sheafer, T., 2013. Social Media and the Arab Spring: Politics Comes First. Int. J. Press. 18, 115–137. https://doi.org/10.1177/1940161212471716

Wu, S., Hofman, J.M., Mason, W.A., Watts, D.J., 2011. Who says what to whom on twitter, in: Proceedings of the 20th International Conference on World Wide Web - WWW ’11. ACM Press, p. 705. https://doi.org/10.1145/1963405.1963504

Yang, J.H., Rojas, H., Wojcieszak, M.E., Aalberg, T., Coen, S., Curran, J., Hayashi, K., Iyengar, S., Jones, P.K., Mazzoleni, G., Soroka, S., Tiffen, R., Papathanassopoulos, S., Rhee, J.W., Rowe, D., Soroka, S., Tiffen, R., 2016. Why Are “Others” So Polarized? Perceived Political Polarization and Media Use in 10 Countries. J. Comput. Commun. 21, 349–367. https://doi.org/10.1111/jcc4.12166

Yang, M., Wen, X., Lin, Y.R., Deng, L., 2017. Quantifying Content Polarization on Twitter, in: 2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC). IEEE, School of Computing and Information, University of Pittsburgh, United States, pp. 299–308. https://doi.org/10.1109/CIC.2017.00047

Yardi, S., Boyd, D., 2010. Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. Bull. Sci. Technol. Soc. 30, 316–327. https://doi.org/10.1177/0270467610380011

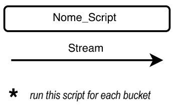
Zhang, W., Seltzer, T., Bichard, S.L., 2013. Two Sides of the Coin: Assessing the Influence of Social Network Site Use During the 2012 U.S. Presidential Campaign. Soc. Sci. Comput. Rev. 31, 542–551. https://doi.org/10.1177/0894439313489962

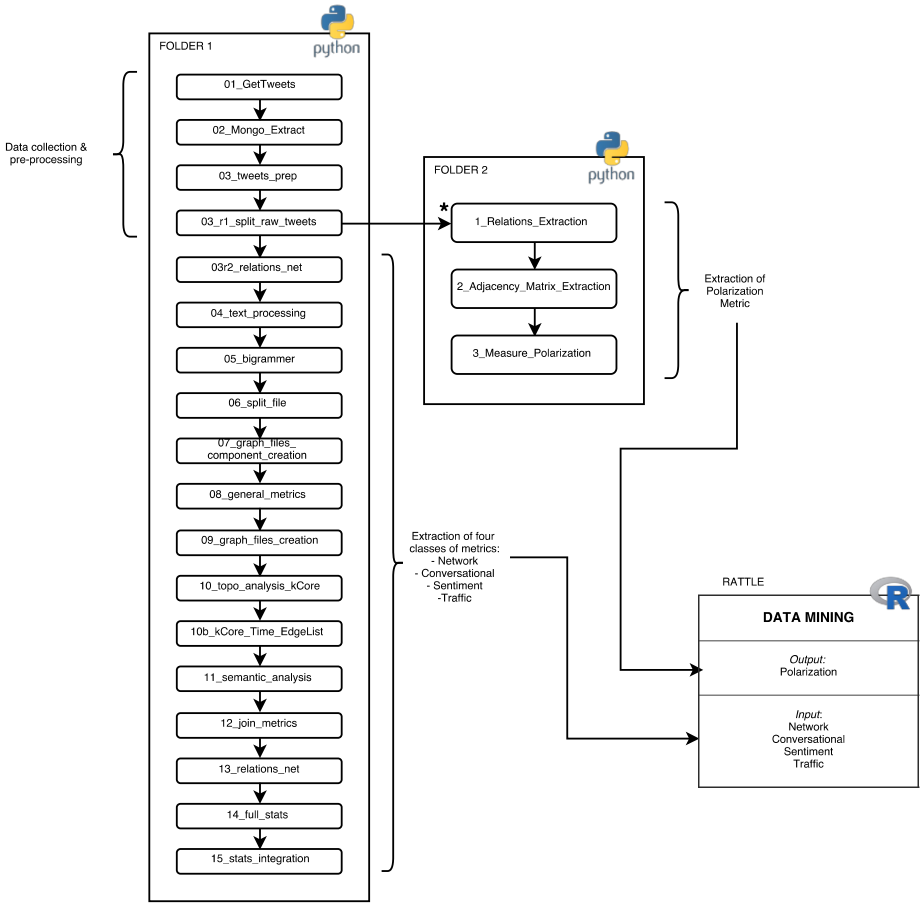
Zhu, Q., Skoric, M., Shen, F., 2017. I Shield Myself From Thee: Selective Avoidance on Social Media During Political Protests. Polit. Commun. ISSN 34, 112–131. https://doi.org/10.1080/10584609.2016.1222471

APPENDIX

Appendix A

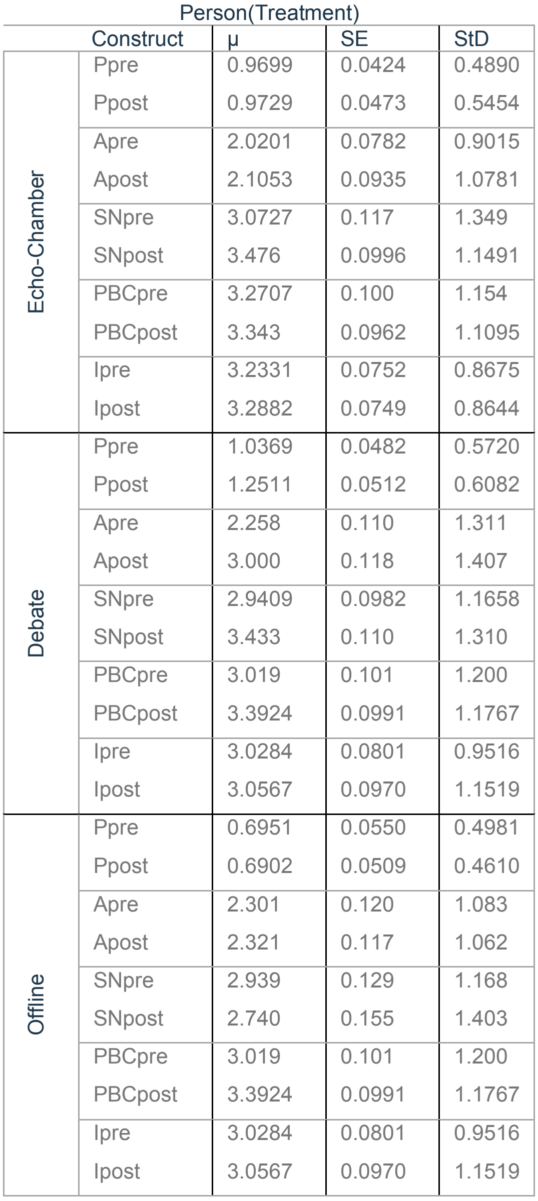
WORKFLOW





Appendix B

Descriptive Statistic



Person(Treatment)				
	Construct	μ	SE	StD
Echo-Chamber	Ppre	0.9699	0.0424	0.4890
	Ppost	0.9729	0.0473	0.5454
	Apre	2.0201	0.0782	0.9015
	Apost	2.1053	0.0935	1.0781
	SNpre	3.0727	0.117	1.349
	SNpost	3.476	0.0996	1.1491
	PBCpre	3.2707	0.100	1.154
	PBCpost	3.343	0.0962	1.1095
	Ipre	3.2331	0.0752	0.8675
	Ipost	3.2882	0.0749	0.8644
Debate	Ppre	1.0369	0.0482	0.5720
	Ppost	1.2511	0.0512	0.6082
	Apre	2.258	0.110	1.311
	Apost	3.000	0.118	1.407
	SNpre	2.9409	0.0982	1.1658
	SNpost	3.433	0.110	1.310
	PBCpre	3.019	0.101	1.200
	PBCpost	3.3924	0.0991	1.1767
	Ipre	3.0284	0.0801	0.9516
	Ipost	3.0567	0.0970	1.1519
Offline	Ppre	0.6951	0.0550	0.4981
	Ppost	0.6902	0.0509	0.4610
	Apre	2.301	0.120	1.083
	Apost	2.321	0.117	1.062
	SNpre	2.939	0.129	1.168
	SNpost	2.740	0.155	1.403
	PBCpre	3.019	0.101	1.200
	PBCpost	3.3924	0.0991	1.1767
	Ipre	3.0284	0.0801	0.9516
	Ipost	3.0567	0.0970	1.1519

			Group(Treatment)						
Construct	Group	μ	SE	StD	Construct	Group	μ	SE	StD
Ppre	1	1,093	0,16	0,618	Ppost	1	1,107	0,135	0,523
	2	0,653	0,0985	0,3815		2	0,987	0,126	0,487
	3	1,05	0,131	0,524		3	1,213	0,138	0,554
	4	1,092	0,11	0,397		4	0,877	0,181	0,651
	5	1,06	0,158	0,499		5	1,1	0,161	0,51
	6	0,909	0,161	0,533		6	0,982	0,165	0,547
	7	0,908	0,115	0,413		7	0,692	0,115	0,413
	8	1,15	0,0732	0,207		8	1	0,193	0,545
	9	0,787	0,116	0,45		9	0,72	0,138	0,533
	10	1,035	0,134	0,553		10	1,082	0,139	0,575
	11	0,253	0,0836	0,3642		11	0,484	0,0814	0,3548
	12	0,464	0,101	0,364		12	1,185	0,114	0,412
	13	1,133	0,129	0,5		13	1,693	0,0753	0,2915
	14	0,625	0,308	0,871		14	1,3	0,248	0,701
	15	0,276	0,0727	0,2815		15	0,44	0,0709	0,2746
	16	0,093	0,0941	0,3993		16	1,189	0,126	0,533
	17	1,436	0,132	0,437		17	1,636	0,132	0,437
	18	1,038	0,0417	0,1668		18	1,45	0,0592	0,2366
	19	0,415	0,0478	0,1725		19	1,8	0	0
	20	0,046	0,0243	0,0877		20	1,8	0	0
Construct	Group	μ	SE	StD	Construct	Group	μ	SE	StD
Apre	1	1,956	0,237	0,916	Apost	1	1,844	0,262	1,015
	2	1,867	0,208	0,805		2	2,489	0,366	1,419
	3	2,042	0,199	0,797		3	2,208	0,274	1,095
	4	1,923	0,221	0,795		4	2,205	0,294	1,059
	5	1,600	0,147	0,466		5	2,30	0,370	1,17
	6	2,212	0,197	0,654		6	2,545	0,306	1,014
	7	2,077	0,239	0,862		7	1,615	0,180	0,650
	8	1,958	0,391	1,105		8	2,25	0,361	1,020
	9	2,489	0,350	1,356		9	2,089	0,295	1,144
	10	1,961	0,225	0,927		10	1,745	0,231	0,954
	11	2,596	0,230	1,004		11	2,982	0,304	1,326
	12	3,385	0,932	3,361		12	2,462	0,455	1,642
	13	1,822	0,186	0,722		13	2,978	0,399	1,545
	14	2,083	0,197	0,556		14	3,167	0,445	1,260
	15	2,200	0,215	0,834		15	2,156	0,335	1,296
	16	2,333	0,222	0,943		16	3,019	0,319	1,355
	17	2,303	0,236	0,781		17	2,364	0,358	1,187
	18	1,917	0,186	0,745		18	3,562	0,362	1,449
	19	1,897	0,221	0,798		19	3,667	0,297	1,072
	20	1,949	0,155	0,559		20	3,615	0,354	1,275

Group(Treatment)

Construct	Group	μ	SE	StD	Construct	Group	μ	SE	StD
SNpre	1	3,356	0,275	1,065	SNpost	1	4,022	0,281	1,087
	2	3,556	0,225	0,87		2	4,822	0,336	1,301
	3	3,083	0,334	1,336		3	3,563	0,354	1,418
	4	3,077	0,344	1,241		4	3,692	0,365	1,316
	5	2,367	0,356	1,127		5	2,900	0,500	1,58
	6	3,000	0,333	1,106		6	3,242	0,365	1,212
	7	3,179	0,371	1,338		7	3,538	0,355	1,28
	8	3,500	0,403	1,141		8	3,625	0,477	1,35
	9	2,711	0,274	1,061		9	3,533	0,399	1,547
	10	2,882	0,262	1,08		10	3,000	0,347	1,429
	11	3,140	0,244	1,062		11	2,281	0,231	1,008
	12	2,154	0,238	0,857		12	3,308	0,334	1,205
	13	3,311	0,314	1,218		13	4,289	0,208	0,805
	14	2,875	0,408	1,154		14	2,250	0,417	1,179
	15	3,133	0,125	0,483		15	4,222	0,347	1,344
	16	3,037	0,295	1,252		16	2,981	0,316	1,341
	17	2,758	0,247	0,818		17	4,364	0,208	0,69
	18	2,708	0,277	1,108		18	4,333	0,215	0,861
	19	2,359	0,315	1,134		19	3,641	0,375	1,35
	20	2,462	0,324	1,167		20	3,846	0,303	1,094

Construct	Group	μ	SE	StD	Construct	Group	μ	SE	StD
PBCpre	1	3,133	0,266	1,03	PBCpost	1	3,200	0,331	1,284
	2	3,333	0,333	1,291		2	3,222	0,344	1,331
	3	3,208	0,317	1,268		3	3,354	0,309	1,235
	4	3,410	0,312	1,123		4	3,590	0,185	0,669
	5	2,967	0,328	1,036		5	3,000	0,479	1,515
	6	3,091	0,331	1,096		6	2,970	0,152	0,505
	7	3,333	0,33	1,191		7	3,820	0,415	1,497
	8	3,083	0,495	1,4		8	3,333	0,244	0,69
	9	3,778	0,245	0,948		9	4,834	0,279	1,082
	10	3,235	0,297	1,223		10	3,235	0,212	0,872
	11	3,737	0,269	1,174		11	2,895	0,249	1,083
	12	3,026	0,369	1,33		12	3,615	0,335	1,208
	13	3,778	0,271	1,048		13	2,733	0,301	1,166
	14	3,125	0,403	1,14		14	3,417	0,603	1,707
	15	3,756	0,273	1,058		15	2,600	0,233	0,902
	16	3,278	0,247	1,049		16	2,741	0,242	1,026
	17	2,727	0,381	1,263		17	4,152	0,217	0,721
	18	2,417	0,286	1,145		18	3,813	0,228	0,911
	19	3,769	0,271	0,976		19	2,359	0,332	1,197
	20	2,667	0,314	1,13		20	3,718	0,306	1,104

Group(Treatment)									
Construct	Group	μ	SE	StD	Construct	Group	μ	SE	StD
Ipre	1	3,356	0,191	0,74	Ipost	1	2,911	0,191	0,74
	2	3,311	0,267	1,035		2	3,333	0,237	0,917
	3	3,292	0,221	0,885		3	3,250	0,229	0,915
	4	3,179	0,205	0,741		4	3,000	0,223	0,805
	5	3,067	0,313	0,991		5	3,067	0,262	0,829
	6	3,242	0,312	1,034		6	3,515	0,208	0,689
	7	3,333	0,233	0,839		7	3,103	0,254	0,917
	8	3,292	0,292	0,825		8	3,625	0,38	1,076
	9	3,178	0,208	0,805		9	3,689	0,226	0,877
	10	3,078	0,237	0,976		10	3,451	0,196	0,807
	11	3,421	0,19	0,83		11	2,491	0,212	0,925
	12	2,641	0,246	0,887		12	2,615	0,232	0,837
	13	2,556	0,212	0,823		13	2,311	0,224	0,868
	14	3,708	0,133	0,375		14	2,833	0,236	0,667
	15	3,978	0,18	0,695		15	1,933	0,17	0,657
	16	3,426	0,217	0,92		16	2,796	0,329	1,396
	17	2,515	0,303	1,004		17	4,515	0,138	0,456
	18	3,042	0,221	0,885		18	4,250	0,127	0,509
	19	2,462	0,205	0,74		19	3,769	0,175	0,629
	20	2,308	0,133	0,48		20	3,564	0,175	0,629