UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DOTTORATO DI RICERCA IN INGEGNERIA DELL'INFORMAZIONE

XXXIV CICLO

Opinion Mining and Clusters Detection in Online Public Debates: a Quantitative Analysis

Alessandro Galeazzi

Relatore:
Prof. Francesco Gringoli, Università degli Studi di Brescia

Correlatore:
Prof. Walter Quattrociocchi, Università degli Studi di Roma "La Sapienza"

Settore Scientifico Disciplinare ING-INF/03

The dissertation of Alessandro Galeazzi is approved.

Advisor:
Prof. Francesco Gringoli, University of Brescia

Co-advisor:
Prof. Walter Quattrociocchi, “La Sapienza” University of Rome

The dissertation of Alessandro Galeazzi has been reviewed by:

Prof. Petra Kralj Novak, Department of Network and Data Science, Central European University, Austria

Prof. Andrea Gabrielli, Department of Engineering, University Roma Tre, Italy

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank Francesco Gringoli and Walter Quattrociocchi for their precious advice, extraordinary commitment and patient guidance throughout my research.

I am thankful to all the people I had the opportunity to work with. Niccolò Di Marco, Carlo Michele Valensise, and Fabiana Zollo deserve a special mention. Gabriele Etta went beyond being an exceptional colleague and became a good friend. I am also profoundly in debt to Matteo Cinelli for being an incredible guide, a wonderful colleague, and, most of all, a sincere friend.

A big thank you to all my friends and teammates, who patiently listened to countless unsolicited and unnecessary explanations of my work. I am grateful to Giovanna for reading and revising all my abstracts and messy sentences. I am in debt to my old friend Chiara, for helping me survive since high school, without asking anything back. A shout-out to ”I Fioi del Negroni” who were always available to uplift me and share good times.

I also thank my parents and brothers, whose encouragement was fundamental in reaching this goal.

Lastly, a very special thought to Francesca. Her absolute trust in my abilities and unconditional support during the bad days of these years contributed to this achievement at least as much as I did.

This thesis would have never been possible without them.

Vita

2016
B.Sc. in Ingegneria dell'informazione
108/110
University of Padova, Padova, Italy

2018
M.Sc. in ICT for Internet and Multimedia
110/110 cum laude
University of Padova, Padova, Italy

2018
M.Sc. in Communication Engineering
National Taiwan University, Taipei, Taiwan

Publications

[1] M. Gentil, A. Galeazzi, F. Chiariotti, M. Polese, A. Zanella, M. Zorzi, "A Deep Neural Network Approach for Customized Prediction of Mobile Devices Discharging Time", *GLOBECOM 2017-2017 IEEE Global Communications Conference*, 1-6

[2] M. Cinelli, S. Cresci, A. Galeazzi, W. Quattrociocchi, M. Tesconi, "The limited reach of fake news on Twitter during 2019 European elections", *PloS one* 15 (6), e0234689

[3] G. Bonaccorsi, F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A.L. Schmidt, C.M. Valensise, A. Scala, W. Quattrociocchi, F. Pammolli, "Economic and social consequences of human mobility restrictions under COVID-19", *Proceedings of the National Academy of Sciences* 117 (27), 15530-15535

[4] P. Galeazzi, A. Galeazzi, "The ecological rationality of decision criteria", *Synthese*, 1-24

[5] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C.M. Valensise, E. Brugnoli, A.L. Schmidt, P. Zola, F. Zollo, A. Scala, "The COVID-19 Social Media Infodemic", *Scientific Reports* 10 (16598 (2020))

[6] M. Cinelli, G.D.F. Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, "The echo chamber effect on social media", *Proceedings of the National Academy of Sciences* 118 (9)

[7] A. Galeazzi, L. Badia, S.-C. Chang and F. Gringoli, "Reputation-Based Spectrum Data Fusion against Falsification Attacks in Cognitive Networks", *19th Mediterranean Communication and Computer Networking Conference (MedComNet)*, 2021

[8] A. Galeazzi, M. Cinelli, G. Bonaccorsi, F. Pierri, A.L. Schmidt, A. Scala, F. Pammolli, W. Quattrociocchi, "Human mobility in response to COVID-19 in France, Italy and UK", *Scientific Reports* 11 (1),1-10

[9] C.M. Valensise, A. Serra, A. Galeazzi, G. Etta, M. Cinelli, W. Quattrociocchi, "Entropy and complexity unveil the landscape of memes evolution", *Scientific Reports* 11, 20022 (2021)

[10] G. Etta, M. Cinelli, A. Galeazzi, C.M. Valensise, M. Conti, W. Quattrociocchi, "News consumption and social media regulations policy", *arXiv preprint, arXiv:2106.03924*(Under Review)

[11] C.M. Valensise, M. Cinelli, M. Nadini, A. Galeazzi, A. Peruzzi, G. Etta, F. Zollo, A. Baronchelli, W. Quattrociocchi, "Lack of evidence for correlation between COVID-19 infodemic and vaccine acceptance", *[Under Review]*

[12] J. Flamino, A. Galeazzi, S. Feldman, M.W. Macy, B. Cross , Z. Zhou, M. Serafino, A. Bovet, H.A. Makse, B.K. Szymanski, "Twitter Influencers and Increased Polarization during two U.S. Presidential Elections", *(Under review)*

Abstracts & Posters

[1] A. Galeazzi, "Human mobility in response to COVID-19 in France, Italy and UK", *Urban Complex Systems 2020*, 2020

Presentations

[1] A. Galeazzi, "Human mobility in response to COVID-19 in France, Italy and UK", *Urban Complex Systems 2020*, December 2020

[2] A. Galeazzi, "Reputation-Based Spectrum Data Fusion against Falsification Attacks in Cognitive Networks", *19th Mediterranean Communication and Computer Networking Conference (MedComNet)*, June 2020

ABSTRACT

The advent of online platforms dramatically changed the way people create and communicate content. In online social media, users can easily share information that thousands of peers may consume almost immediately. Moreover, the unique features offered by online social platforms also allow immediate feedback and interactions, creating the perfect environment for the proliferation of an intense debate around controversial topics. Nevertheless, this new and disintermediated type of communication and platforms' feed algorithms may influence the dynamics of online discussion, creating a fertile environment for the formation of clusters of users reinforcing their opinion through repeated interactions called echo chambers.

In this thesis, we study the debate around controversial topics in online social media, such as political elections and disease outbreaks, and analyze the factors influencing its dynamics. We also assess the impact of unsubstantiated rumors and measure the shift in polarization around political elections. Finally, we compare the effect of echo chambers around several topics and across different social

media and quantify the online infodemic concurrent with the recent pandemic. In our studies, we find evidence that users tend to cluster together into groups with opposite opinions around debated topics and consume information adhering to their system of beliefs. This characteristic appears to dominate the information consumption dynamics in online social media, influencing the spread of both confirmed news and unsubstantiated rumors.

SOMMARIO

La nascita di piattaforme online tramite cui condividere informazioni con una platea virtualmente illimitata ha radicalmente cambiato il modo di comunicare. Attraverso i social media, chiunque è in grado di creare contenuti che non solo sono fruiti quasi in tempo reale da migliaia di utenti, ma che, grazie alle funzioni offerte dalle varie piattaforme online, possono ottenere un feedback immediato tramite commenti e reazioni. Questa modalita' di comunicazione veloce e disintermediata, da un lato, fornisce il mezzo perfetto per la proliferazione di dibattiti su temi controversi, dall'altro, grazie anche alla presenza di algoritmi che riducono la diversita' dei contenuti a cui un utente è esposto, crea l'ambiente perfetto per la formazione di gruppi ideologicamente omogenei di persone, definiti echo chambers. In questi ambienti, grazie a ripetute interazioni con altri ideologicamente affini, gli utenti sono esposti a una visione parziale e omogenea dell'argomento dibattuto, che li porta a rinforzare la propria opinione preesistente e ignorare posizioni contrarie.

Questa tesi si pone l'obiettivo di analizzare molteplici aspetti che influenzano

il dibattito online. In particolare, è stata studiata l'evoluzione del dibattito nei social media riguardo argomenti controversi quali elezioni politiche e pandemia, evoluzione della polarizzazione e impatto di notizie non verificate sulle elezioni presidenziali americane, la presenza di echo chambers in varie piattaforme e attorno diversi argomenti di dibattito; è stata inoltre misurata la magnitudo dell' "infodemia" concomitante con la recentente pandemia. Lo studio dimostra come gli utenti, quando dibattono sui social media attorno ad argomenti controversi, tendono ad aggregarsi in fazioni ideologicamente opposte, consumando informazioni che rinforzano la loro visione e ignorando altri punti di vista. Questa caratteristica sembra dominare il consumo di informazioni online, influenzando la diffusione sia di contenuti fondati sia di informazioni non verificate.

CHAPTER 1

INTRODUCTION

Online social media platforms allow us to create and communicate content to an incredible amount of people and these interactions are the source of a huge amount of information about several users' characteristics. Thus, these digital traces can be exploited to study people's behavior in online social media and this new perspective attracted many researchers from various disciplines proposing different approaches to enhance this emerging field.

In this thesis, we use several techniques to perform quantitative analyses of social media data for studying different aspects of users' behavior on online platforms. In particular, we rely on complex network theory to highlight interaction structures among users, we use a source-based approach to estimate misinformation circulation, we perform statistical analysis on information diffusion patterns and use epidemiological models to mimic the news spreading dynamic, and we employ artificial intelligence techniques to perform semantic analysis on text. The scope of this work is to understand the role that several factors have in the diffusion of information, ranging from the characteristic of the content itself to the topological aspect of the diffusion network and the impact of the automated accounts and suggestion algorithms.

1.1 State of the Art

The advent of the internet has radically changed many aspects of our daily life. Nowadays, people use online resources in the most diverse areas of their life, from work to entertainment. One of the most renewed aspects is how users access, consume and interact with information. In the past, the information flow, and thus the agenda setting, was mainly unidirectional from the source (such as newspapers or tv programs) to the audience. Now, people are not just exposed to a plethora of different information sources but can also produce, share and communicate their original contents in a disintermediated, or at least algorithm mediated, way. Throughout online social media, users can create and share contents potentially with the entire society. Moreover, the unique features implemented in such platforms allow users to interact among themselves and receive immediate feedback on what they publish.

This new way of producing and sharing content dramatically changed the paradigm of accessing and consuming information: it democratized the information market potentially making every user also a publisher. However, this change in the information consumption dynamic brought new potential threats to the public society. In particular, disinformation and misinformation, that is, the spread of inaccurate or false information, has become a problem of primary interest to the point that the World Economic Forum listed it as one of the most critical threats of our society [1].

Thus, the spread of fake news has attracted the attention of many researchers from various fields and several aspects of rumors spreading dynamics and online news consumption have been investigated. In particular, one thread of research focused on the data-driven modeling of factors influencing the spreading of true and false news. In [2] authors studied the spreading of rumors on Facebook and showed that polarization is one of the main drivers of news spreading. Authors of [3] studied the news consumption pattern on Facebook showing that users tend to interact with a limited set of pages, while Cinelli et al. investigated the users' news consumption and find that selective exposure, i.e. the tendency of users to acquire only information that adheres to their beliefs, dominates the Facebook news consumption patters [4]. Also, authors of [5] studied the news consuming behavior of 10 million of Facebook users and found that the exposure to ideologically diverse content is limited by personal choices more than algorithmic factors.

Together with the analysis of users' news consumption patterns, which highlighted the role of human behaviors such as selective exposure and selection bias in limiting the exposition to opposite opinions, researchers also studied the topological structures that influence the spread of information and misinformation.

Echo Chambers. One of the most popular concepts in this area is the idea of echo chambers [2], [6]–[10]. Intuitively, an echo chamber can be defined as a group of users that interact among themselves on a debated topic and reinforce their opinion due to repeated contacts with others sharing the same beliefs. Recently, both the existence and the impact on the news spreading dynamics of echo chambers have been questioned [11]–[13] revealing that the debate about such a structure and its possible effect on news consumption dynamics is still open in the academic community.

Social Bots. Another aspect that has attracted the attention of researchers is the influence of automated accounts on the spread of true and false news. Authors of [14] compared the spread of information and misinformation in Twitter and found that automated accounts (bot) play a limited role, while in [15] authors claim that bots play a substantial role in the spread of low credibility content. These opposite results may derive from the different definitions of what misinformation is and show that more research has to be done to clarify online spreading dynamics [16].

Misinformation and Society. All the attention academics dedicated in studying misinformation and the conditions that may foster its spreading is justified by the importance this phenomenon has on the public society. Indeed, the influence of social media in various aspects of the public society such as political election, behavior adoption, and so on, has been not clarified yet and concerns about the possible negative effects on the democratic process have been raised. For example, researchers studied the influence of fake news and bots on the 2016 U.S. presidential election on Twitter and found that bots and misinformation may have harmed the online debate [17]. However, Bovet et al. found that users' propensity to consume fake news may depend on their political leaning [18] and authors of [19] showed that misinformation is consumed mainly by a small portion of users with well-defined characteristics. Nevertheless, the concern about the negative influence of social media on the outcome of political elections has been so high that in 2019 Twitter announced it would not allow any political advertisement on the platform [20]. The influence of fake news and bots on the public debate has been under the attention of governments and academia also in many other circumstances, such as Brexit and 2019 European elections [21]–[23]. More recently, the concern about the impact of misinformation about COVID-19 has been so high that WHO immediately launched an online platform to mitigate the phenomenon [24].

The spread of misinformation and its consequences are not the only threats that online social media could bring to public society, but also the fragmentation of the online ecosystem into echo chambers with opposite views has been a reason of concern for the public society [25]. Indeed, the idea that online social media may foster polarization and increase the ideological distance between opposite factions has gained momentum due to events such as the Pittsburgh synagogue shooting or the Christchurch mosque shooting [26]. For example, in the first case, the author of the attack was very active on Gab, a far-right social media that performs little control on the content posted [27], and the same platform has been considered responsible for the dissemination of the video of the Christchurch attack[28]. More recently, social media have been suspected to have fostered the capitol hill riot [29], [30], where some Trump supporters occupied the United States Capitol to stop the formalization of Joe Biden's victory. Part of the public opinion considered this event as the consequence of the tweets and declaration made by Trump in the months before the formalization date in which he sustained the presence of electoral fraud and denied to admit Joe Biden victory [31]. Trump was also formally impeached for "incitement of insurrection" by the U.S. Congress, moreover Facebook and Twitter banned his account because "the risks of allowing President Trump to continue to use our service during this period are simply too great" [32] and "due to the risk of further incitement of violence" [33]. This unprecedented decision on the one hand triggered the protests against "Big Tech censorship" [34], on the other hand, highlighted the lack of understanding and regulation of the interplay between online social media and offline public society.

1.2 Advances

In this thesis, we analyze the online debate around controversial themes of public interest, such as political elections or COVID-19 outbreaks, quantifying the presence and the impact of misinformation on online discussion. We also study the effect of news feed algorithms on polarization and how they may affect the spreading of information and misinformation. We rely on data taken from mainstream social media such as Twitter, Facebook, and Youtube as well as less know platforms such as Gab. The idea of using digital traces to study users' online behavior can be traced back to [35], in which the concept of computational social science is introduced. One advantage of this approach is that it allows researchers to rely on an amount of data that was not available using other research methods such as surveys or lab experiments. Moreover, some psychology and social sciences studies that exploited surveys or lab experiments as research technique may suffer from reproducibility issue [36], [37]. Recently, through the analysis of social media data, researches obtained relevant insights on various aspects of online human behavior, such as bounds on the users' maximum number of contacts [38], news diet and selection bias [4], [5], polarization around debated topics and controversial events[3], [18], [39]–[41], and echo chambers [42]–[44]. Nevertheless, the debate in the academic community is still open and several aspects of users' online behavior are still to be clarified. For instance, most of these studies focused only on one platform, and hence their result may not hold for other social media. Second, the environment itself is subject to changes over time: social platforms may occasionally modify their feed algorithms, their policy on the contents allowed, or their features. Although caution should be taken when generalizing the knowledge obtained from social media data, research exploiting this methodology is providing us important insights on events, such as political elections or COVID-19 outbreak, in almost real-time.

In this thesis, we present a series of works addressing different aspects of online social media. First, we study the interaction patterns of different public figures during the 2019 European elections, showing the limited circulation fake news had during the electoral campaign. Second, we analyze the difference in the debate around the 2016 and 2020 U.S. presidential elections on Twitter, assessing the role of automated accounts and quantifying the differences in terms of polarization. We continue our study with a comparative analysis of the polarization and the echo chamber effect on four social media around several debated topics, highlighting clues for the feed algorithms effect on users' segregation. Finally, we analyze the dramatic increase in the amount of news about COVID-19 shared at the beginning of the first virus outbreak, focusing on the differences and similarities in the spread dynamic of information and misinformation.

Our research shows that users tend to cluster into groups with homogeneous beliefs around debated topics. Such a structure, called echo chambers, may be fostered by different factors, one of which is the platform feed algorithms. Indeed, a stronger feed algorithm may enhance the segregation of users according to their beliefs and reduce the exposure to opposite contents, fostering the echo chamber effect. Moreover, we found that online misinformation tends to circulate among specific clusters of users and its volume is far lower with respect to reliable information. In general, most users rarely interact with misinformation sources, but some strongly endorse this kind of content and are very prone to share them. This is coherent with the presence of echo chambers, in some of which misinformation may easily proliferate. Nevertheless, the limited reach of fake news may be also the consequence of the policies adopted by online platforms to limit the spread of such contents. Finally, our results about the COVID-19 debate show that misinformation and reliable information have comparable spreading patterns. Although the share of misinformation circulating changes from platform to platform, the dynamic diffusion is the same for both reliable and questionable contents.

The rest of this thesis is structured as follow: Chapter 2, 3, 4 and 5 are dedicated to the works about 2019 European Elections, U.S. 2020 presidential election, echo chamber effect, and COVID-19 infodemic respectively. In chapter 5 conclusions and future work are presented, while in Appendices A,B and C reports tables and statistics of the dataset used in the studies.

CHAPTER 2
METHODS AND DATASETS

The works presented in this thesis required several techniques to gather and analyze the datasets. In this section, we present an overview of the methods used for downloading, processing, and analyzing the data.

2.1 Data Download

This section is dedicated to summarizing the techniques used for gathering data from different platforms, which can vary from the use of official API to scraping. However, to be compliant with the legislation and the terms of services only publicly available information has been downloaded from each platform.

2.1.1 Twitter

Twitter provides an official API documented at `https://developer.twitter.com/en/docs` to retrieve different contents. Their API allows searching for tweets matching specific keywords, users details, follower-following relationships, and engagement information. However, Twitter applies limitations in terms of both the rate limit and the amount of information that can be downloaded.

2.1.2 Reddit

Reddit provides an official API that allows access to several types of publicly available content, such as submissions and subreddit data, user information, and engagement details. Moreover, the website `https://pushshift.io/` provides a near real-time mirroring of all public available Reddit content, accessible via a registration-free API.

2.1.3 Facebook

Facebook provided the Facebook Graph API to access their publicly available data. Together with data on public available posts, this API allowed to retrieve also user-level information, such as likes on posts, replies, and so on. Currently, this API has been dismissed.

2.1.4 Gab

Gab does not have any officially documented API, but it is still possible to download data directly querying the website. However, the rate limit bounds severely impact the amount of information accessible. To overcome this problem, we developed a multi IP parallel crawler to stream all the contents in near real-time. Moreover, a dataset containing one year of Gab post is available at the website `https://pushshift.io/`.

2.1.5 Instagram

Since no official API are available for Instagram data, we built our process to collect public content related to our keywords and we manually took notes of posts, comments to populate the Instagram Dataset.

2.1.6 Youtube

Youtube data can be collected with the official API documented at `https://developers.google.com/youtube/v3`. This API provides access to different contents such as channel information, comments under videos, and video statistics. Although it is not possible to directly search for videos containing one or more keywords in the title or description, an endpoint that returns a list of video *related* to specific keywords is provided.

2.2 Quantitative Techniques

In this section, we provide an overview of the various techniques we used to analyze the data. To overcome the heterogeneous challenges faced in our research, we adopt several methods from different sectors such as computer science, complex networks theory, statistics, epidemiology, and machine learning.

2.2.1 Network Representation

In our work, we frequently use network representation to represent interaction patterns among entities. We build directed and undirected networks to characterize relationships among users, while we rely on bipartite networks to represent relationships between users and other entities.

Directed networks are built using, for instance, retweets, follower-following, or other interaction information. In such cases, nodes represent users and edges represent interactions. Edges' direction can vary depending on the type of the considered relationship and the purpose of the study.

Undirected networks are used to represent symmetric relationships such as co-commenting or co-liking. In these networks, nodes represent users, and two users are connected if and only if they performed the same action, for example commenting under the same post. In such cases, the direction of an edge is not meaningful nor possible to assess.

Bipartite networks are used to represent interactions between different types of entities, for example between users and posts or users and domains. In these cases, the two entities are represented by two different types of nodes and edge representing relationships between entities can connect only nodes of two different types.

2.2.2 Similarity Network Analysis

In our work, we use similarity network analysis to identify clusters of influential users based on the audience they reach. For each user i in the set of I=$\{1, \ldots, i_n\}$, called the set of influencers, we create a vector $\vec{S^i}$ of size U, where U is the size of the set of unique users in our dataset. Given a user $u \in U$ and and influencers i, the element s_u^i of vector $\vec{S^i}$ defines the number of times user u has interacted with influencer i. We define the adjacency matrix $\mathbf{A}$ of size $I \times I$ for our similarity networks by setting $a_{i,j}$ to the cosine similarity between vectors $\vec{S^i}$ and $\vec{S^j}$. It follows that the higher the cosine similarity, the more users have the similar number of interactions for influencers i, j. An undirected weighted network is created using the adjacency matrix $\mathbf{A}$ and communities are detected by the Louvain algorithm [45].

We quantify the severity of community splitting using two measures of separation between communities. The first is modularity, which computes the sum of the difference between the fraction of edges within each community and such fraction expected within this community in a random network with the same number of nodes and edges. This metric has a range of $[-0.5, 1]$ [46]. A positive value indicates the presence of communities separated from each other, and the closer the modularity is to 1, the stronger communities are separated.

The second measure uses the normalized cut, which is the sum of the weights of every edge that links a pair of communities divided by the sum of the weights of all edges. The result has a range of $[0, 1]$ where the smaller the value, the stronger the separation among communities.

2.2.3 Latent Ideology

The latent ideology estimation allows inferring the ideological position of users around a debated topic and is based on users' interactions such as likes, retweets, mentions, comments, and so on. It follows the method developed in [12], [47], and we use correspondence analysis [48] (CA) to infer users' ideological positions, as done in [12].

Let $\mathbf{A}$ be the adjacency matrix of the network between a set of users called influencers and all the users that interacted with them. The element a_{ij} of $\mathbf{A}$ is equal to the number of times user i interacted with influencer j.

The CA method is executed in the following steps [49]. The matrix of standardized residuals of the adjacency matrix is computed as $\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{rc})\mathbf{D}_c^{-1/2}$, where $\mathbf{P} = \mathbf{A}(\sum_{ij} a_{ij})^{-1}$ is the adjacency matrix normalized by the total number of interactions, $\mathbf{r} = \mathbf{P1}$ is the vector of row sums, $\mathbf{c} = \mathbf{1}^T\mathbf{P}$ is the vector of column sums, $\mathbf{D}_r = \text{diag}(\mathbf{r})$ and $\mathbf{D}_c = \text{diag}(\mathbf{c})$. Using the standardized residuals allows the inference to account for the variation of popularity and activity of the influencers and the users, respectively [12]. Then, a SVD is computed such that

$\mathbf{S} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^T$ with $\mathbf{U}\mathbf{U}^T = \mathbf{V}\mathbf{V}^T = \mathbf{I}$ and $\mathbf{D}_\alpha$ being a diagonal matrix with the singular values on its diagonal. The positions of the users are given by the standard row coordinates: $\mathbf{X} = \mathbf{D}_r^{-1/2}\mathbf{U}$ where we only consider the first dimension, corresponding to the largest singular value. Finally, the ideological positions of the users are found by standardizing the row coordinates to have a mean of zero and a standard deviation of one. The ideological position of the influencers is given by the median of the weighted positions of their audience.

2.2.4 User's Leaning

In our work, one of the most frequently required tasks is the quantification of users' leaning around debated topics through the analysis of the content they produced. We measure the *individual leaning* of a user i toward a specific topic x_i through the analysis of the content it produced. Let a_i be the number of contents produced by user i and define $\mathcal{C}_i = \{c_1, c_2, \ldots, c_{a_i}\}$ as the set of all contents produced by i. The individual leaning of user i can be defined as the average of the leanings of produced contents,

$$x_i \equiv \frac{\sum_{j=1}^{a_i} c_j}{a_i}. \tag{2.1}$$

2.2.5 Homophily in the Interaction Network

Given a network of users, homophily can be defined as the nodes' tendency to interact with others with similar characteristics. In network terms, this translates into a node i with a given leaning x_i more likely to be connected with nodes with a leaning close to x_i [10]. In our studies, this concept can be quantified by defining, for each user i, the average leaning of their neighborhood, as:

$$x_i^N \equiv \frac{1}{k_i^{\rightarrow}} \sum_j A_{ij} x_j \tag{2.2}$$

where A_{ij} is the adjacency matrix of the interaction network, $A_{ij} = 1$ if there is a link from node i to node j, $A_{ij} = 0$ otherwise, and $k_i^{\rightarrow} = \sum_j A_{ij}$ is the out-degree of node i. The presence of homophily is assessed by studying the relationship $x_i \sim x_i^N$.

2.2.6 Dynamic Information Spreading Model

To model the spreading dynamics of news and rumors, we leverage on the SIR model. In the SIR model, each agent can be susceptible, infectious, or recovered. Susceptible users may become infectious upon contact with infected neighbors, with a specific transmission probability β, and infectious users can recover with probability ν. We run the SIR dynamics on the interaction networks of users starting the epidemic process with only one node i infected and stopping it when no more infectious nodes are left.

The set of nodes in a recovered state at the end of the dynamics started with user i as a seed of infection, i.e., those that become aware of the information initially propagated by user i forms the *set of influence* of user i, $\mathcal{I}_i$ [50]. We define the average leaning of the set of influence of user i, μ_i, as

$$\mu_i \equiv |\mathcal{I}_i|^{-1} \sum_{j \in \mathcal{I}_i} x_j. \tag{2.3}$$

To analyze the impact of users' leaning on information spreading, we define the average leaning $\langle \mu(x) \rangle$ of the influence sets reached by users with leaning x and study the relationships $\langle \mu(x) \rangle \sim x$.

2.2.7 Topic Modeling

In our study, we apply topic modeling to analyze the context around online debates. We build word embeddings for the text corpus and then we cluster words by the Partitioning Around Medoids (PAM) algorithm on their vector representations to assess the topics around which the perception of the COVID-19 debate is concentrated.

Our word embeddings are the distributed representations of words learned by neural networks in which each word is represented as a vector in $\mathbf{R}^n$. This representation has the property that similar words are embedded into vectors close to each other. Skip-gram model [51] is used to construct word embedding for each corpus. Given the sequence of words $w_1, w_2, \ldots, w_T$ that is the representation of a content, stochastic gradient descent with gradient computed through backpropagation rule [52] is employed to maximize the average log probability

$$\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{j=-k}^{k}\log p(w_{t+j}|w_t)\right] \tag{2.4}$$

where k is the size of the training window.

The Skip-gram model associates an input and output vectors u_w and v_w to each word w.Thus, we define the probability of correctly predicting the word w_i given the word w_j as:

$$p(w_i|w_j) = \frac{\exp\left(u_{w_i}^T v_{w_j}\right)}{\sum_{l=1}^{V}\exp\left(u_l^T v_{w_j}\right)} \tag{2.5}$$

where V is the number of words in the corpus vocabulary. The training quality is affected mainly by two parameters, namely the dimensionality of word vectors(i.e. the dimension of the word embedding representation space) and the size of the surrounding words window

To increase the performance, preprocessing steps in which we remove special strings such as urls, stop-words, and other special charters and patterns are performed.

To identify topics, clustering by using the Partition Around Medioids (PAM) technique with the cosine distance matrix of words in the vector representation as proximity metric is performed. The average silhouette width for different values of the number of clusters k is calculated to select the best values and the cluster stability is evaluated by computing the average pairwise Jaccard similarity between clusters based on 90% sub-samples of the data. Lastly, we produce word clouds to identify the topic of each words cluster.

2.2.8 Epidemiological Models

In our studies, we use two epidemiological models to estimate the infodemic growth, namely the exponential model of [53] (EXP) and the classical SIR model [54] (SIR). The exponential model provides an estimate of the basic reproduction number R_0 and has been successfully employed in data-scarce settings and shown to be on-par with more traditional compartmental models.

The following equation fully describes the model:

$$I = \left[\frac{R_0}{(1+d)^t}\right]^t \tag{2.6}$$

where I is the incidence, t is the number of days, R_0 is the basic reproduction number, and d is a damping factor accounting for the transmissibility reduction over time. The other model that is used in our work is the classical SIR model [54]. In this model, a susceptible population can be infected with a rate β by coming into contact with infected individuals, but infected individuals can recover with a rate ν. The following set of differential equations describe the model:

$$\begin{aligned} \partial_t S &= -\beta S \cdot I/N \\ \partial_t I &= \beta S \cdot I/N - \nu I \\ \partial_t R &= \nu I \end{aligned} \tag{2.7}$$

where S is the number of susceptible, I is the number of infected and R is the number of recovered.

$R_0 = \beta/\nu$, also known as the basic reproduction number, corresponds to the proportion between the rate of infection β and the rate of recovery ν.

The basic reproduction numbers R_0^{EXP} and R_0^{SIR} for the EXP and the SIR model are estimated by using least-square estimates of the model parameters[55].

2.3 Data Processing Tools

This section is dedicated to the description of the tools developed and utilized to manipulate our datasets.

2.3.1 URL Classification

One common task that is frequently performed on datasets is URL classification. Relying on third-party data, we build lists of domains related to the online news outlet that contains information about the level of reliability and other feature of each source. Thus, from each URL we extract the domains, and then each domain is classified according to the information in our list if present.

2.3.2 Short URL Resolver

Frequently, URLs are shortened using shortening URL services such as Bitly or Cuttly. To extract the domain from such type of URL is first necessary to retrieve the unshortened URL. To resolve shortened URLs, we rely on a tool that sends head requests over the internet for each shortened URL and stores the answer. Given the considerable size of our datasets and in order to speed up the process, a multi-IP parallel approach has been adopted.

2.3.3 Account Geolocation

On online platforms, users may occasionally leave hints about the place they live and this information can be used to assign them a location.

To perform this task, we exploit several different geolocator services such as Google Maps, Bing, and GeoNames that offer their services via Web APIs. Thus, we identify geographical hints and used them to query one of the available geolocators and associated the corresponding geographic coordinates to the corresponding account.

CHAPTER 3

QUANTIFYING THE IMPACT OF MISINFORMATION AND BOTS ON TWITTER 2019 EUROPEAN ELECTIONS ONLINE DEBATE

Fake news and misinformation have been listed to be a major threat to our society and concerns about the possible impact of fake news on the outcome of political election have been raised. Moreover, automated account are suspected of fake news dissemination to influence election results. In this chapter, we analyze the impact of fake news on the 2019 European Elections on Twitter. In particular, we followed the activity of accounts owned by users with different roles in the public society such as politicians, news outlets, and show business personalities. We considered also recognized fake news spreader and analyze the interactions with other actors, providing important insight on the role and the impact of fake news during this election.

3.1 Introduction

The wide diffusion of online social media platforms such as Facebook and Twitter raised concerns about the quality of the information accessed by users and about the way in which users interact with each other [2], [39], [40], [56]–[60]. Recently, the chairman of Twitter announced that political advertisements will be banned from Twitter soon, claiming that our democratic systems are not prepared to deal with the negative consequences brought by the power and influence of online advertising campaigns [20]. In this context, a wide body of scientific literature focused on the influence and on the impact of disinformation and automation (i.e., social bots) on political elections [14], [17], [18], [41], [61]–[65]. In [18] the authors studied the impact of fake news on the 2016 US Presidential elections, finding that users sensitivity to misinformation is linked to their political leaning. In [61] is highlighted that fake news consumption is limited to a very small fraction of users with well defined characteristics (middle aged, conservative leaning and strongly engaged with political news). Authors of [14] studied the spreading of news on Twitter in a 10 years time span and found that, although false news spread faster and broader than true news, social bots boost false and true news diffusion at the same rate. The pervasive role of social bots in the spread of disinformation was instead reported in [66] for financial discussions, where as much as 71% of users discussing hot US stocks were found to be bots. The effects of fake and unsubstantiated news affected also the outcome of other important events at international level. For instance, the evolution of the Brexit debate on Facebook has been addressed in [67] where evidence about the effects of echo chambers, confirmation bias in news consumption and clustering are underlined. Nevertheless, as stated in [16], the conclusions of these and other studies are partially conflicting. This conflict can be the result of the differences in the definitions of fake news or misinformation adopted by different authors, that have somewhat contributed in switching the attention from the identification of fake news to the definition itself.

In particular, authors in [2] and [3] focused their attention on the process that can boost the spreading of information over social media. In these works, it is highlighted that phenomena such as selective exposure, confirmation bias and the presence of echo chambers play a pivotal role in information diffusion and are able to shape the content diet of users. Given the central role of echo chambers in the diffusion process, authors of [68] propose a methodology based on users polarization for the early identification of topics that could be used for creating misinformation. However, in [69] it is stressed that the phenomenon of echo chambers can drive the spreading of misinformation and that apparently there are no simple ways to limit this problem.

Considering the increasing attention paid to the influence of social media on the evolution of the political debate, it becomes of primary interest to understand, at a fast pace, how different actors participate in the online debate. Such concerns were renewed in the view of the US Presidential Election of November 2020 or the future national elections in EU countries.

The goal of our work is to characterize the information flow among different actors that took place in the run up to the last European Parliament elections held between the 23rd and 26th of May, 2019. According to the European legislation, every 5 years all the country members of the EU have to hold elections to renovate their members at the European Parliament. The election can be held in a temporal window of few days and every state can decide in which days to hold the voting procedure. During the electoral campaign, concerns about the impact of fake news on the upcoming European election were risen by several news outlets [21] and misinformation have been monitored, also thanks to the effort of NGOs, in different platforms [22]. The EU itself started a joint and coordinated action on misinformation mitigation [70]. Based on what happened during Brexit and the US 2016 election, also EU leaders encouraged the adoption of measures at the European level to counteract the diffusion and impact of Fake News [23]. Additional evidence of the potential impact of misinformation during European Election motivated studies at national level such as [71]. Starting from these premises, our study aims to assess the reach of fake news during European Elections. In this context, we characterize the public debate on Twitter in the months before the elections. In particular, we aim at understanding which role was played by users that have different positions in public society, including disinformation outlets and popular actors either directly or indirectly related to politics, to obtain a wide view of the process. Through a thorough quantitative analysis on a dataset of 399,982 tweets posted by 863 accounts in the three months before the elections, we first analyze the information flow from a geographical point of view and then we characterize the interactions among different classes of actors. Finally, we compare the impact of disinformation-related accounts with respect to all others. We find that all classes, except official news outlets, have a strong tendency towards intra-class interaction and that the debate rarely cross the national borders. Moreover, disinformation spreaders have a marginal role in the information exchange and are ignored by other actors, despite their repeated attempts to join the conversation. Although the maximum outreach of fake news accounts is lower than that of other categories, when we take into account comparable levels of popularity we observe an outreach for disinformation that is larger than that of traditional outlets and comparable to that of politicians. Such evidence demonstrates that disinformation outlets have a rather active followers base. However, the lack of interactions between fake news accounts and others demonstrated that their user base is confined to a peripheral portion of the network, suggesting that the countermeasures taken by Twitter, such as suspension or ban of suspicious accounts, might have been effective in keeping the Twittersphere clean.

	class	users	tweets	interactions *retweets*	*replies*	*mentions*	*articles*
				retweets	*replies*	*mentions*	*articles*
●	fake	45	24,331	4,375	2,640	12,927	4,389
●	official	333	207,171	49,515	9,966	99,595	48,095
●	politicians	328	88,627	23,188	5,603	57,512	2,324
●	showbiz	98	29,873	5,414	2,838	21,475	146
●	social media	8	8,824	402	3,901	4,499	22
●	sport	37	33,616	6,057	2,059	25,490	10
●	trademarks	6	4,289	207	1,789	2,293	0
●	VIPs	11	3,251	192	812	2,238	9
	total	863	399,982	89,350	29,608	226,029	54,995

Table 3.1: Dataset summary.

3.2 Results and Discussion

By exploiting Twitter APIs, we collected data from the Twitter timelines of 863 users. This resulted in the acquisition of 399,982 tweets shared between February 28 and May 22, 2019. The 863 users in our dataset are classified into 8 categories, based on their roles in the society. In detail, we have categories encompassing trusted news outlets (labeled `official`), politicians, disinformation outlets (`fake`), show business personalities (`showbiz`), official accounts of social media platforms, sport personalities, famous brands (`trademarks`), and other VIPs. By leveraging information contained in tweets and users metadata that we collected, we also computed the interactions between all the accounts of our dataset and we geolocated Twitter users, whenever possible. A detailed view of our Twitter dataset is summarized in Table 3.1 while additional information is available in the "Materials and Methods" Section.

By leveraging account interactions, we built a directed graph $G = (V, E)$ where each node $v_i \in V$ corresponds to a Twitter account and each link $e_i = (v_A, v_B) \in E$ from node v_A to node v_B exists $\big($i.e., Ⓐ$\longrightarrow$Ⓑ$\big)$ if and only if v_A interacted with v_B in one of the following ways: (i) v_A **retweeted** v_B; (ii) v_A **replied** to v_B; (iii) v_A **mentioned** v_B in a tweet; (iv) v_A **tweeted a link to an article** that mentioned v_B. We refer to the last type of interaction as *indirect* – whereas all others are *direct* – since Web links do not point directly to Twitter accounts, but rather point to Web pages outside Twitter that, in turn, mention accounts in our dataset. Our rich interaction network is thus representative of the information flow across different actors, including disinformation outlets, and several countries involved in the 2019 European Parliament elections.

We first characterize the geographical composition of our dataset. As shown in Figure 3.1, our dataset is mainly made up of accounts located in the EU and the US. However, a small fraction of accounts belong to other parts of the world. This is due to the fact that we integrated our initial set of accounts with a subset of popular accounts (more than 1M followers) that interacted with them. Notably, only a small fraction of accounts belong to non EU/US places. This may be a first signal that the interactions rarely cross national borders. Indeed, the top panel of Figure 3.2 shows the geographic distribution of user interactions on a world

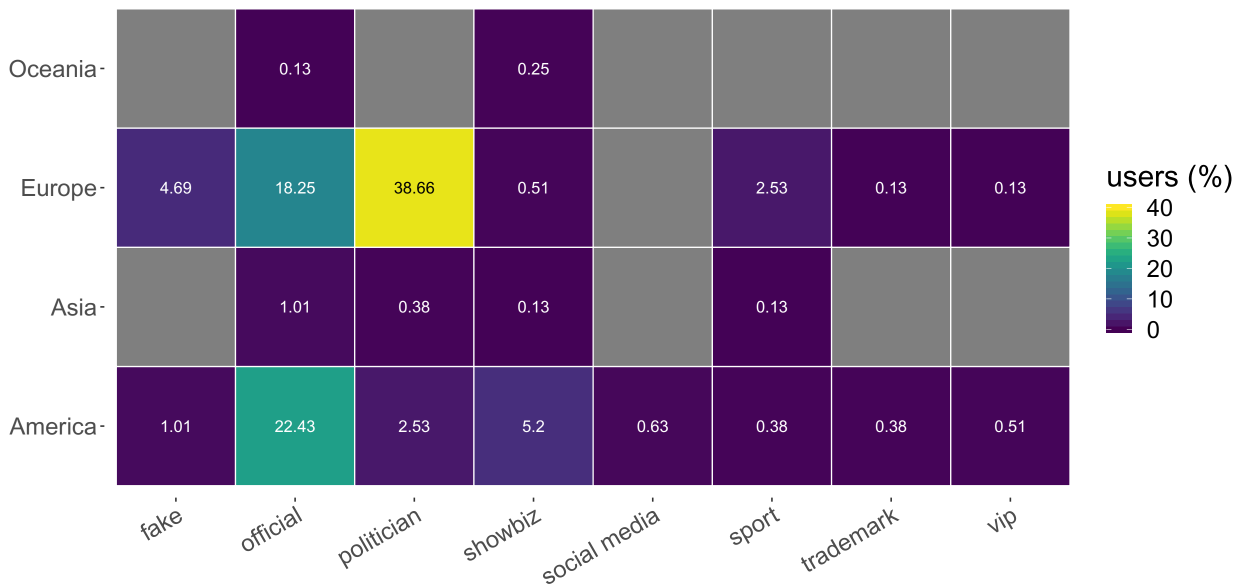


Fig. 3.1: Heatmap showing the distribution of users interacting with the different actor classes, per geographic area.

map, while the bottom panel represents the information as a chord diagram where interactions are grouped by actor class and by country. The top panel highlights that the vast majority of interactions (65%) is initiated by official accounts (green links) and that a considerable number of links between the US and the EU (10%) exists. The chord diagram of Figure 3.2 provides more details about countries and classes, confirming that the biggest contribution to the debate is provided by official accounts, followed by politicians. However, it is noticeable that most of

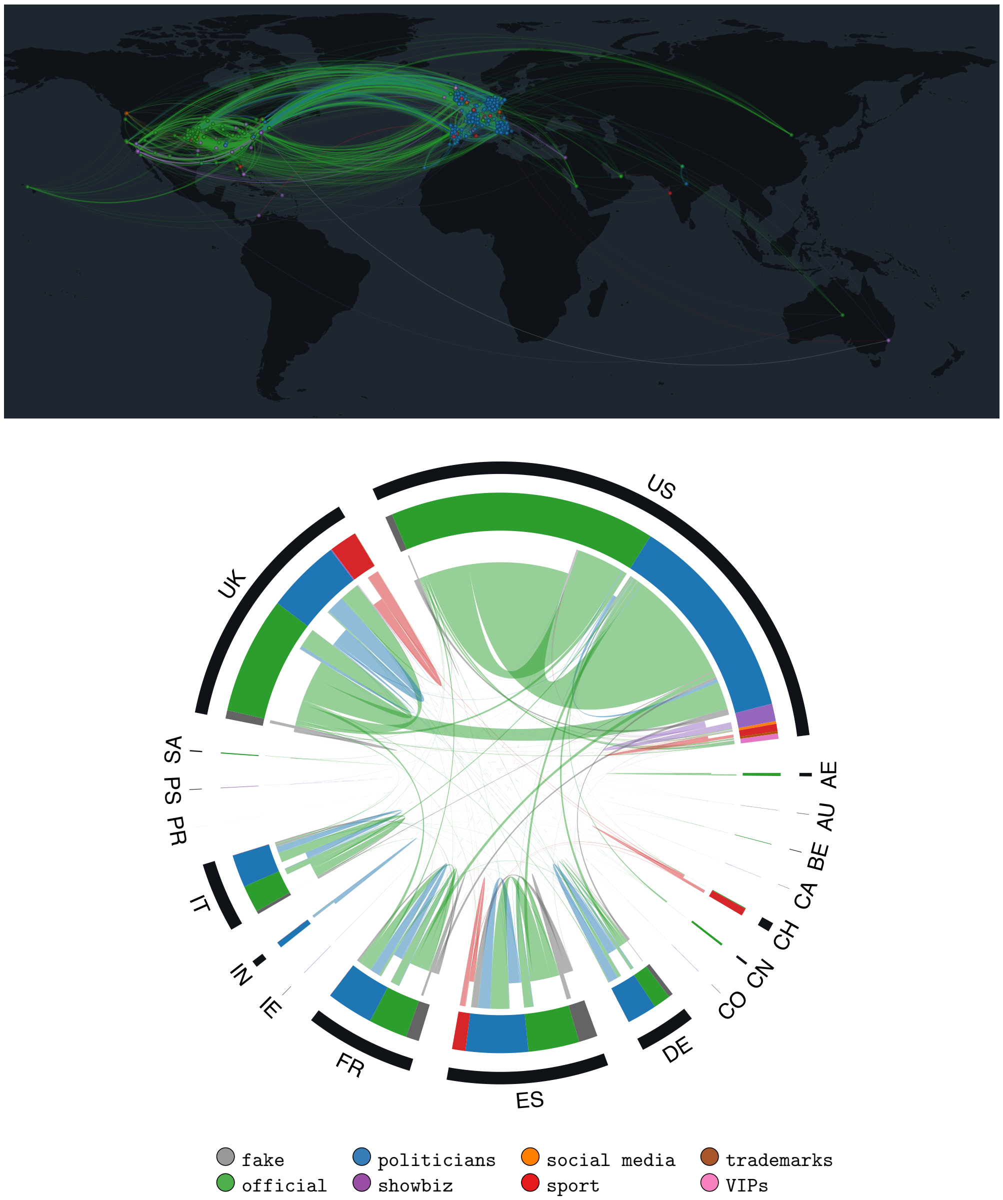


Fig. 3.2: Node-link diagram showing the geographic representation of information flows (top) and Chord diagram showing class interactions grouped by country (bottom) during EU elections. Loops are taken into account only in the chord diagram, that highlights the tendency of accounts to interact mainly with users in the same nation and often also in the same class.

the links start and end in the same country, while the center of the chord diagram is almost empty, implying that the debate rarely crossed national borders. The only relevant exception is represented by official news outlets that tend to cite politicians from other countries (11% of all links). This is particularly clear for the UK, where a relevant fraction of links coming from official accounts point to US politicians (36% of all links from UK news outlets) – that is, UK news outlets tweet about US politicians quite often. All other groups tend to refer only to accounts from the same country and often also of the same type. Although a precise assessment of the causes of this phenomenon is beyond the scope of this work, we provide further details and briefly discuss the possible impact of language barriers in the following paragraph.

country	total	home lang	other langs	home lang from outside	home total ratio	other home ratio	outside total ratio
UK	249,923	243,342	6,581	323,857	0.9737	0.0270	1.2958
USA	246,760	241,735	5,025	325,464	0.9796	0.0208	1.3190
Spain	178,815	159,091	19,724	7,286	0.8897	0.1240	0.0407
France	178,593	152,482	26,111	5,943	0.8538	0.1712	0.0333
Italy	159,814	135,128	24,686	1,024	0.8455	0.1827	0.0064
Germany	135,481	96,187	39,294	967	0.7100	0.4085	0.0071

Table 3.2: The column **country** refers to the geolocation of the tweets at national level. The column **total** is the total number of tweets located in the respective country. The column **home lang** reports the number of tweets using the national language of the country. The column **home lang from outside** reports the number of tweets made in different countries that the national language of the country belonging to the respective row. The other columns report the ratio deriving from the previous columns.

In order to give more details about the impact of language barriers on the debate among nations, we provide more details about the languages used in the six

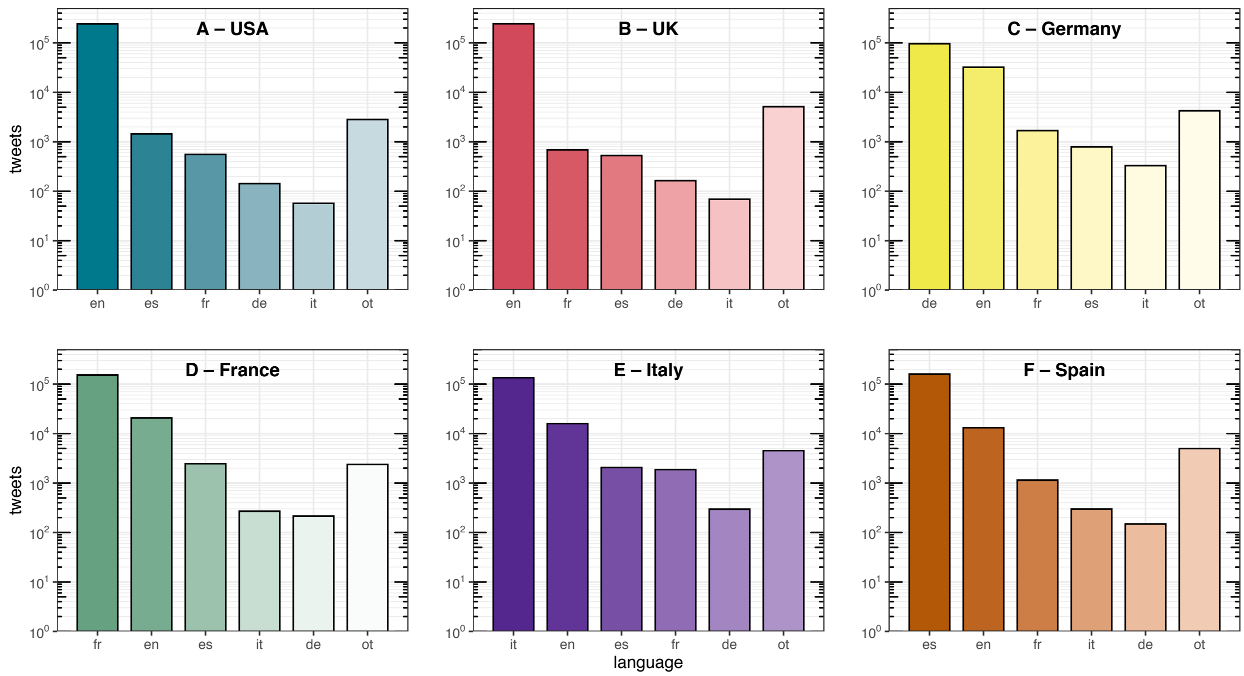


Fig. 3.3: Per country distribution of languages in the tweets of our dataset.

more relevant nations in our analysis. Figure 3.3 shows that the national language is indeed the most frequent in the tweets of the respective country; nonetheless other languages, and especially English in non-English speaking countries, is well represented within the Twittersphere. Table 3.2 provides more details concerning the usage of national and foreign languages in different countries. In summary, we can conclude that despite the national language is the most used in each countries also other languages are quite well represented and therefore the impact of language barriers is present but it is not the unique element that determines the lack of inter-connections among countries.

In order to understand how accounts of the same type interact among them-

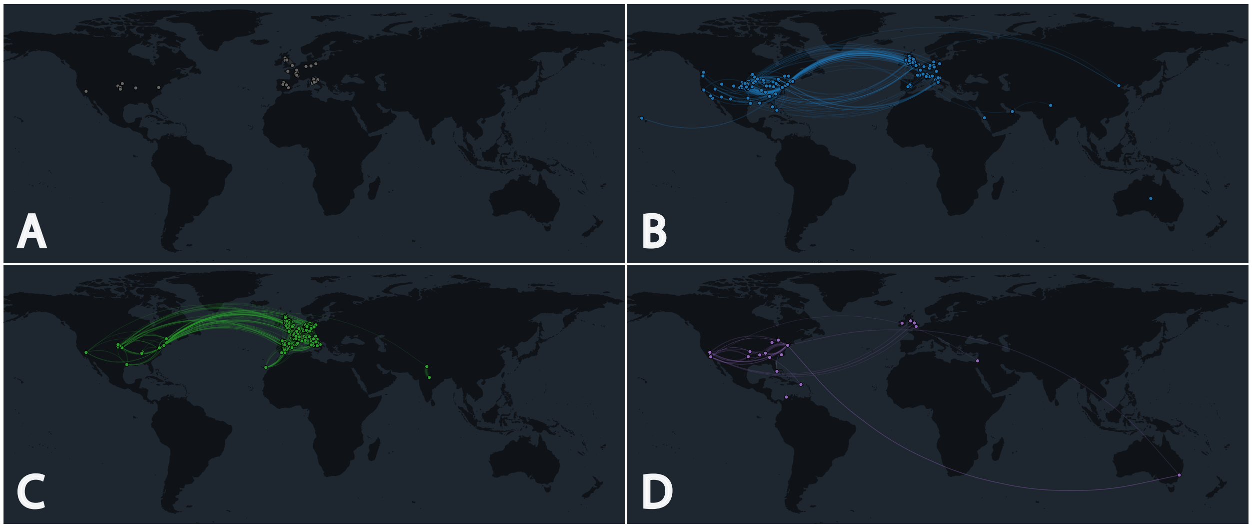


Fig. 3.4: Geographic representation of intra-class interactions for the four biggest classes of actors: fake (panel **A**), politicians (panel **B**), official (panel **C**), and showbiz (panel **D**). Notably, panel **A** has only one link between two nodes in the UK, while all other panels exhibit a large number of interactions. For clarity, self-loops are omitted.

selves, we induced subgraphs based on node categories hence obtaining one subgraph for each category. Figure 3.4 shows the subgraphs plotted in a world map, for the four biggest classes of actors: fake (panel **A**), politicians (panel **B**), official (panel **C**), and showbiz (panel **D**). We note that only subgraphs related to official news outlets, politicians and showbiz accounts are well connected. Indeed, the proportion of nodes belonging to the largest connected component is respectively 66%, 91% and 84% of the total number of nodes. On the contrary, the graph related to disinformation news outlets (panel **A**) comprises mostly isolated nodes. In the case of disinformation news outlets the nodes belonging to the largest connected component are about 9% of the total number of nodes. Such evidence suggests that Twitter accounts related to disinformation outlets rarely dialogue with their peers, but rather they prefer to interact with other types of actors. Furthermore, comparing Figure 3.4 with the chord diagram of Figure 3.2, we can infer that outlets labelled as fake also display a tendency towards self-mentioning. Instead, politicians and showbiz accounts show a relevant percentage of interactions with others from the same class (respectively 42% and 22%, without considering self interactions) while official news outlets interact mainly with other classes (71% of total links amount). Although there are some similarities in the statistics of fake and official outlets – that is, both try to interact with other classes – only official accounts catch the attention of other actors, while fake outlets are most of the times ignored.

To clarify the way in which different actors participated in the debate, we also analyzed the proportion of incoming and outgoing links by class. Results are shown in Figure 3.5. In the first row all types of interactions were considered (i.e., both *direct* and *indirect*), while in the second one only *direct* interactions (retweets, replies, mentions) were taken into account. Some differences arise when comparing all outgoing links with *direct* outgoing links (left-hand side of Figure 3.5), in particular with regards to the classes fake and official. When all kinds of interactions (*direct*+*indirect*) are taken into account, we note an increment in the fraction of outgoing links that point to politicians (blue-colored bar, +57% and +51% re-

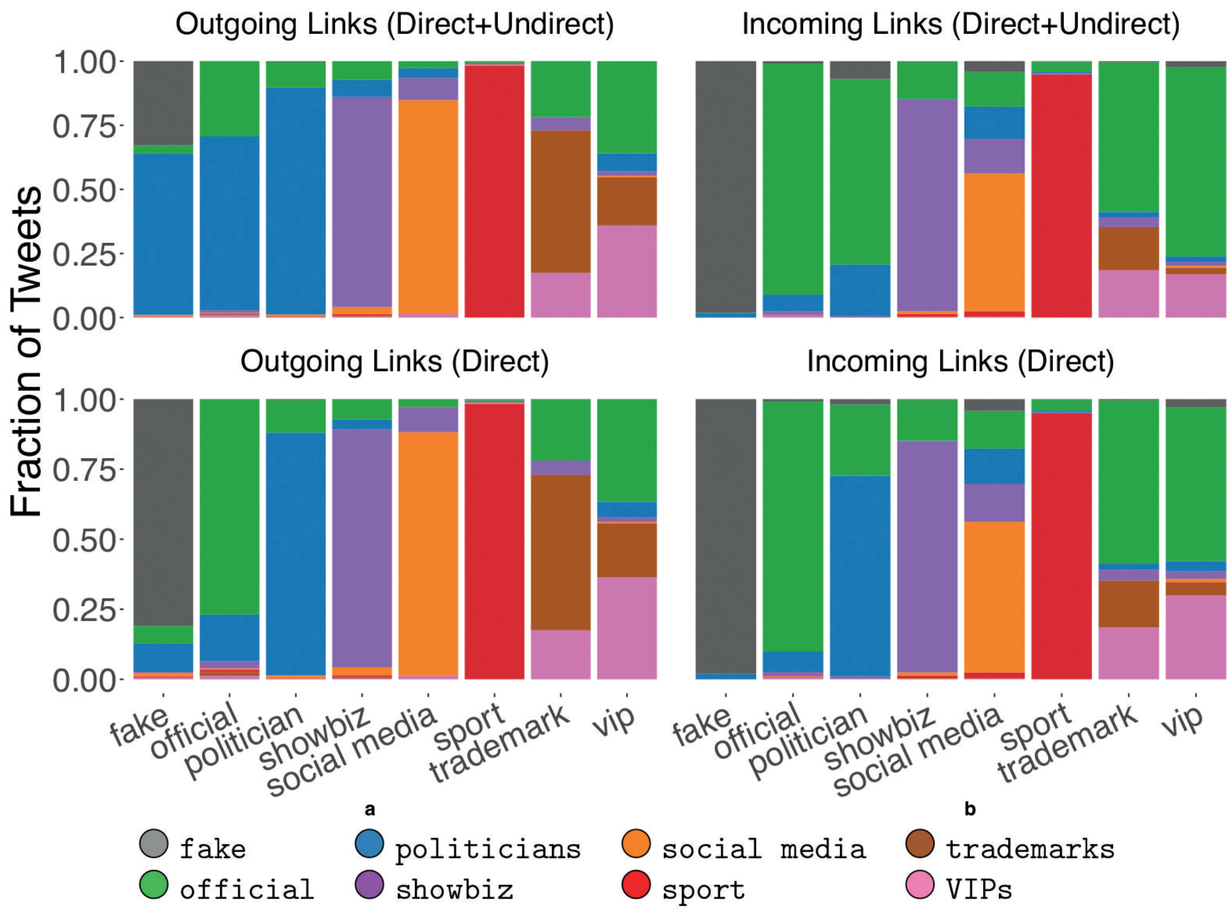


Fig. 3.5: Outgoing and incoming links by class. The top row accounts for all types of interactions, the bottom one only considers *direct* interaction (i.e., replies, retweets, mentions). For this analysis self-loops are considered, which explains the tendency of all classes towards self-interaction.

spectively) for both classes. In other words, the classes labelled as fake and official interact with politicians mainly through external resources. These could be news articles mentioning politicians, that are linked and shared in Twitter.

The proportions of incoming links are shown in the right-hand side of Figure 3.5. The most relevant difference between *direct* and *indirect* links concerns the category of politicians. In fact, there is an increase of links coming from the official and fake classes (+49% and +5%) that is in accordance with the differences found in the case of outgoing links. Again, we notice that accounts, except for official ones, display the tendency to interact mainly within their own classes, and this is even more evident when only *direct* links are taken into consideration.

Finally, by analyzing the behavior of the official and fake classes, we noticed that both of them mainly refer to politicians when external sources are taken into account. However, politicians mostly interact among themselves and only a small fraction (9%) of their outgoing links are directed to official accounts, with disinformation outlets being substantially ignored. Indeed, we measure the number of nodes connected by reciprocal *direct* links (i.e. A and B are connected in both direction with a link representing a mention, a retweet or a reply) among the classes. We found that fake news accounts, news outlets and politicians reach progressively higher reciprocity scores especially within their own classes. The average percentage of nodes connected by reciprocated links in the same class is $\mu = 23.4\%$ and only 9% of fake news accounts are reciprocally interconnected. Moreover, fake news accounts exhibit a behavior that differs from other classes when the percentage of nodes connected by reciprocated inter-class links is taken into account: while the average percentage is $\mu = 5.5\%$, fake news accounts do not display mutual connections with any other class. Such evidence, combined with the information conveyed in Figures 3.2 and 3.4, suggests that disinformation outlets try to fit in the political debate, but they are essentially ignored by mainstream news sources, by politicians, and also by the other classes of actors. Interestingly,the behavior of fake news accounts is akin to that of automated accounts as shown by the authors of [72] during the Catalan Referendum: both fake news and automated accounts tend to target popular users in order to increase their relevance and impact in the public debate.

Our previous finding indicates that Twitter accounts related to disinformation outlets did not seem to be able to enter the main electoral debate. However, despite not attracting interest from the main actors involved in the debate, they could still have had an impact on the general audience. To investigate this issue we study the engagement obtained by the different classes of actors. In particular, each actor produces tweets, and each tweet obtains a certain engagement that derives from the interactions (e.g., retweets) of other users with that tweet. We can thus aggregate the engagement obtained by all tweets of a given actor, to have an indication of the engagement obtained by that actor. Similarly, we can aggregate the engagement obtained by all tweets from actors of a given class (e.g., politicians, fake news outlets, etc.), to have an indication of the engagement obtained by that class of actors. In our study, engagement obtained by a given tweet is simply computed as the number of retweets that tweet obtained. With respect to other measures of engagement (e.g., the number of likes/favorites to a tweet), retweets provide an indication for how much a message spread. As such, they arguably represent a good indicator for investigating the *reach* of fake and authoritative news, which is the goal of our study. Figure 3.6 compares the distribution of the engagement generated by all tweets of disinformation outlets (grey-colored), with those generated by tweets of all the other classes. Overall, Figure 3.6 shows that the engagement obtained by disinformation outlets is lower than that obtained by all other classes. In other words, tweets from accounts in the fake class, tend to receive less retweets than those obtained by other accounts. To dig deeper into this issue, we also considered the popularity of the accounts belonging to the different classes of actors. As a measure of popularity for an account, we considered its number of followers. Then, we compared the relation between the popularity of our accounts and the mean engagement they obtain, for the different classes of actors. Results are shown in Figure 3.7 by means of a bi-dimensional kernel density estimation, for the 6 biggest classes of actors. When we consider also the popularity of the accounts, an important feature of disinformation outlets emerges. Indeed, for mid-low levels of popularity (number of followers $\leq$ 100,000) accounts linked to the spread of disinformation actually obtain more engagement than official news outlets, and almost the same engagement obtained by politicians. This finding is also shown in Figure 3.8, where popularity is logarithmically quantized into 7 buckets. This important finding suggests that the audience of disinformation accounts is more active and more prone to share contents, with respect to that of the other classes. Anyway, no disinformation outlet currently reaches high levels of popularity (number of followers $\geq$ 1M), in contrast with all other classes of actors. As a consequence, highly popular official news outlets still obtain more engagement than disinformation outlets. This indicates that, although the audience of disinformation outlets is more prone to share information than others, their influence on the public debate remains rather limited. Additionally, even though disinformation accounts make efforts to attract interest of other central users, they cannot really fit into the information flow in any significant way.

3.3 Conclusions

We analyzed the interactions of several accounts belonging to different figures of the public society in the context of the 2019 European Parliament elections. To have a wide view of the phenomenon, we included in our dataset also personalities not directly related to politics, such as show business and sport figures, together with a set of disinformation outlets. We collected all the tweets made by the selected accounts in the three months before the election days and we performed a quantitative analysis to identify the characteristics of the debate. By leveraging a semi-automated geolocalization technique, we also performed a geographical analysis of the phenomenon. Results show that the debate on Twitter rarely crossed national borders – that is, accounts tended to interact mainly with others coming from the same nation. Moreover, there was a strong tendency of intra-class interaction – that is, accounts mainly mentioned others from the same class. The only relevant exception were accounts of official news outlets, especially those located in the United Kingdom, that had a non-negligible percentage of links pointing to the US. Moreover, it is interesting that disinformation outlets did not interact among themselves, but rather they exhibited a tendency towards self-mentions and they tried to catch the attention of other popular accounts. Nevertheless, differently from official news outlets, disinformation outlets were almost completely ignored by other actors, thus holding a peripheral position in the interaction network and having a limited influence on the information flow. Still, they exhibited an outreach on general public higher than official news outlet and comparable with the politicians at the same levels of popularity, thus implying that the user base of disinformation outlets was more active than that of other classes of actors. However, all other categories overcame disinformation outlets in terms of absolute maximum outreach, thanks to their significantly larger absolute popularity. Finally, the limited and bounded contribution that disinformation outlets had on the overall interactions suggests that the strategies employed by Twitter to counteract the spreading of disinformation – that is, the ban or suspension of suspicious accounts – may have had a mitigation effect on the spreading of fake news thus preserving the integrity of the Twittersphere.

3.4 Materials and Methods

3.4.1 Data Collection

Our dataset is based on a list of 863 Twitter accounts, split across 8 categories and 18 countries. A pseudonymized version of our dataset is publicly available on GitHub(`https://github.com/cinhelli/Limited-Reach-Fake-News-Twitter-2019-EU-Elections`). Initially, we only considered in our study the 5 biggest European countries (UK, Germany, France, Italy and Spain) and the US. Then, other countries were added when we extended the dataset to also include popular users that interacted with users from our initial set.

The first category of accounts (labeled `fake`) in our study is related to known disinformation outlets. It contains 49 Twitter accounts responsible for sharing disinformation, identified in authoritative reports – such as Reuters' Digital News Report 2018 [73] and a report from the European Journalism Observatory [74] – and fact-checking Web sites – such as Snopes [75] and Politifact [76]. Our list of official news outlets (labeled `official`) contains 347 Twitter accounts. It includes accounts corresponding to the main news outlets in each of the considered countries, derived from the media list released by the European Media Monitor [77], as well as the Twitter accounts of the main US news outlets. We then considered a list of 349 politicians (labeled `politicians`). This list includes all available Twitter accounts of the members of the European parliament [78] as of March 2019, as well as the main politicians for each considered country that did not belong to the European parliament.

We firstly exploited Twitter APIs to crawl the timelines of all the accounts belonging to the 3 previous lists. In order to match the electoral period, we only retained tweets shared between February 28 and May 22, 2019. After this step, we also manually classified a small subset of popular users (more than 1M followers) that interacted with those of our initial list in the considered time period. These accounts were classified in 5 additional categories, based on their role in the society. In this way, we obtained additional 100 `showbiz` accounts (e.g., actors, tv hosts, singers), 10 `social media` accounts (e.g., Youtube's official account), 37 `sport` accounts (e.g., sport players and the official accounts of renown sport teams), 6 `trademarks` accounts related to famous brands (e.g., Nike, Adidas) and 11 accounts of `VIPs` (e.g., the Pope, Elon Musk, J.K. Rowling). For each of these additional accounts, we crawled the respective timeline and only retained tweets shared in our considered time period.

After this data collection process, we ended up with the dataset summarized in Table 3.1, comprising more than 850 labeled accounts and almost 400,000 tweets.

3.4.2 Account Interactions

For each account, we also computed its interactions with other accounts. In particular, we split interactions into 4 different categories: retweets, replies, mentions, and article mentions.

The first 3 types of interactions are straightforward, while an article mention is detected when an account shares a tweet containing a URL to a Web page that mentions one of the labeled accounts in our dataset. To obtain information about article mentions we scraped all the Web pages linked within the tweets of our dataset. Within each page, we performed language detection and named entity recognition. Finally, we cross-checked person named entities with our lists of users.

3.4.3 Account Geolocation

Whenever possible, we also exploited the *location* field of Twitter accounts (both the 863 labeled ones, as well as all others with which they interacted) in order to geolocate them.

For this process, we exploited several different geolocators (e.g., Google Maps, Bing, GeoNames) that offer their services via Web APIs. We first selected all accounts with a non-empty *location* field. Then, we built a blacklist for discarding those locations that were too vague or clearly ironic (e.g., global, worldwide, Mars, the internet), as is frequently the case with user-generated input. For each distinct location that was not removed during the filtering step, we queried one of the available geolocators and we associated the corresponding geographic coordinates to all accounts with that location.

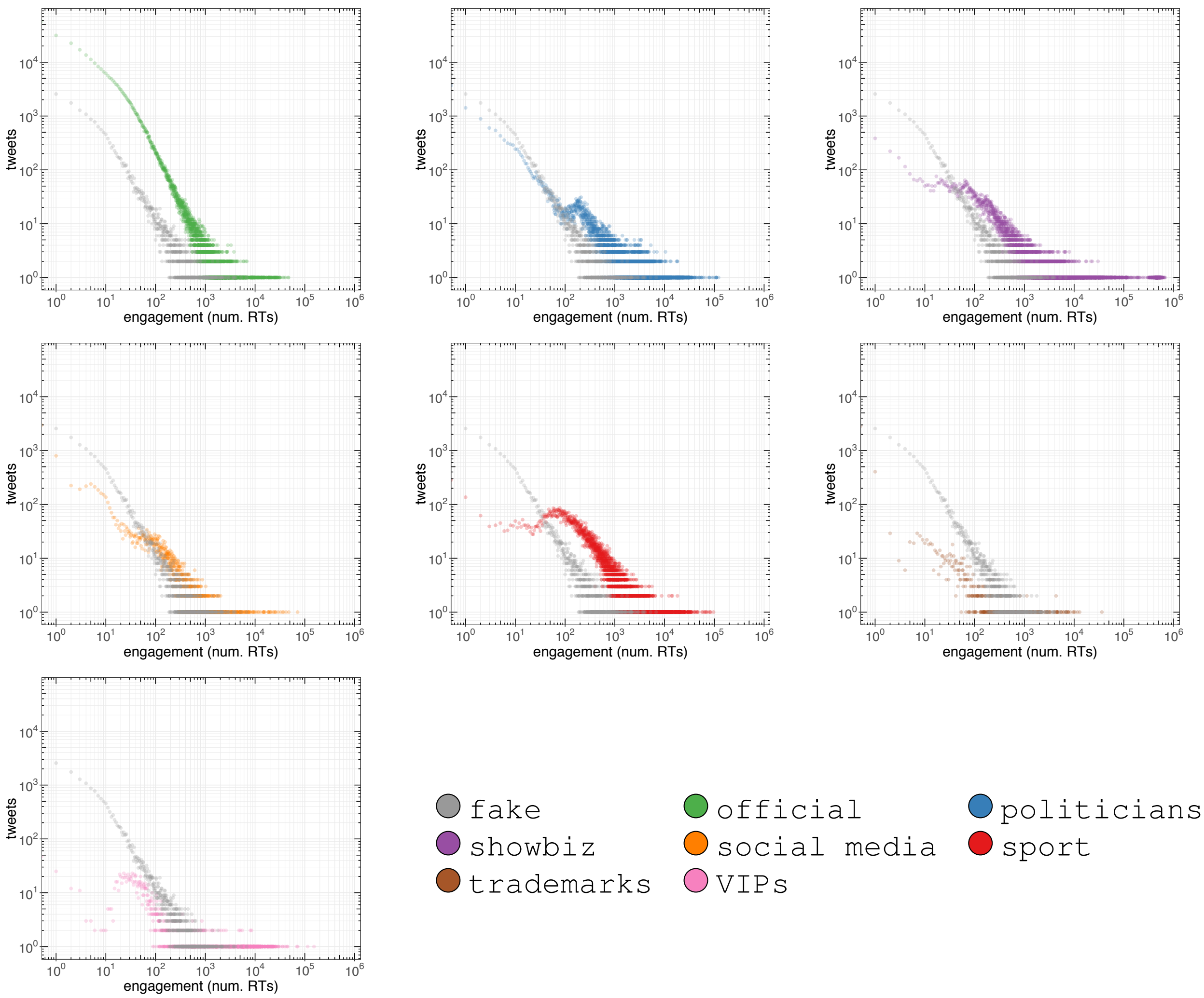


Fig. 3.6: Distribution of the engagement obtained by tweets of disinformation outlets (grey-colored) and comparison with the engagement obtained by tweets of all other classes. Overall, disinformation outlets obtain less engagement than others, as shown by their distribution spanning smaller values on the x axis. Engagement for a given tweet is computed as the number of retweets obtained by that tweet.

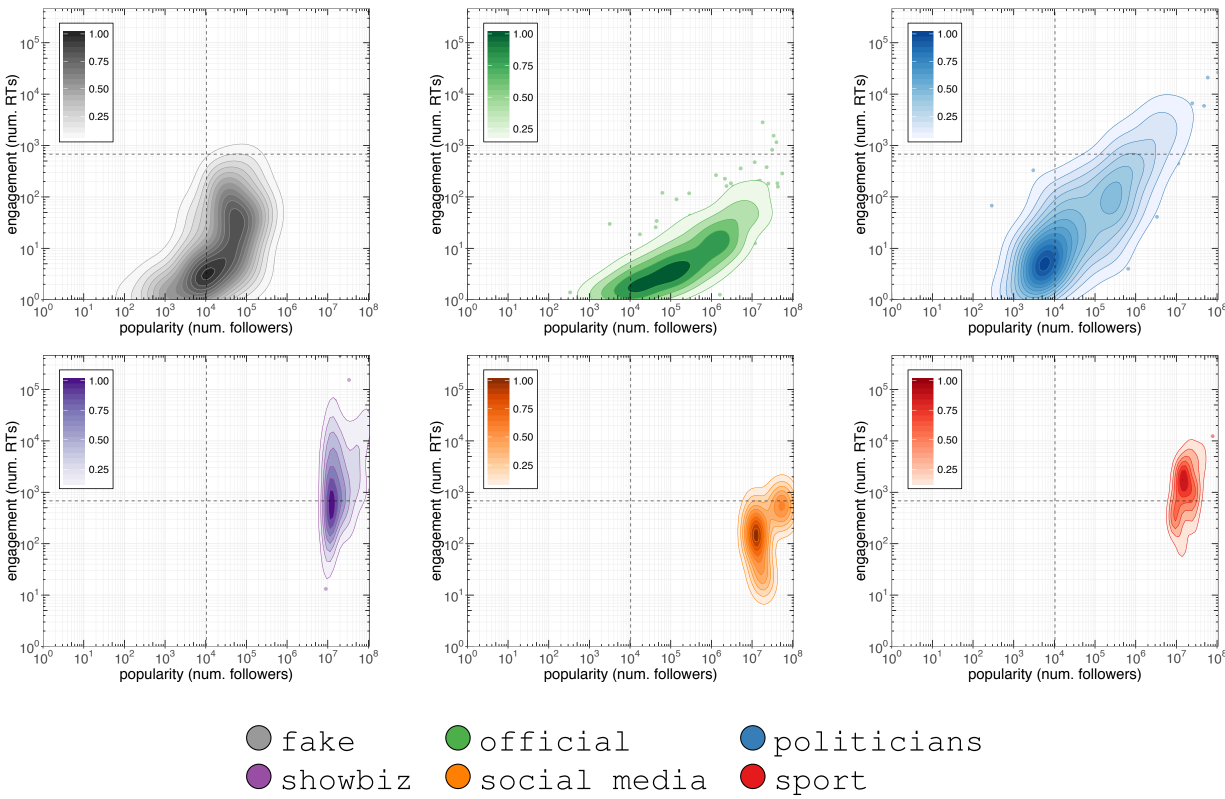


Fig. 3.7: Kernel density estimation of engagement and popularity of the accounts belonging to the main classes of actors. Despite obtaining overall less engagement, disinformation outlets (grey-colored) actually obtain more engagement than official news outlets (green-colored) at middle and low popularity levels. Popularity for a given user is computed as its number of followers. Engagement for a given user is computed as the mean number of retweets obtained by tweets of that user.

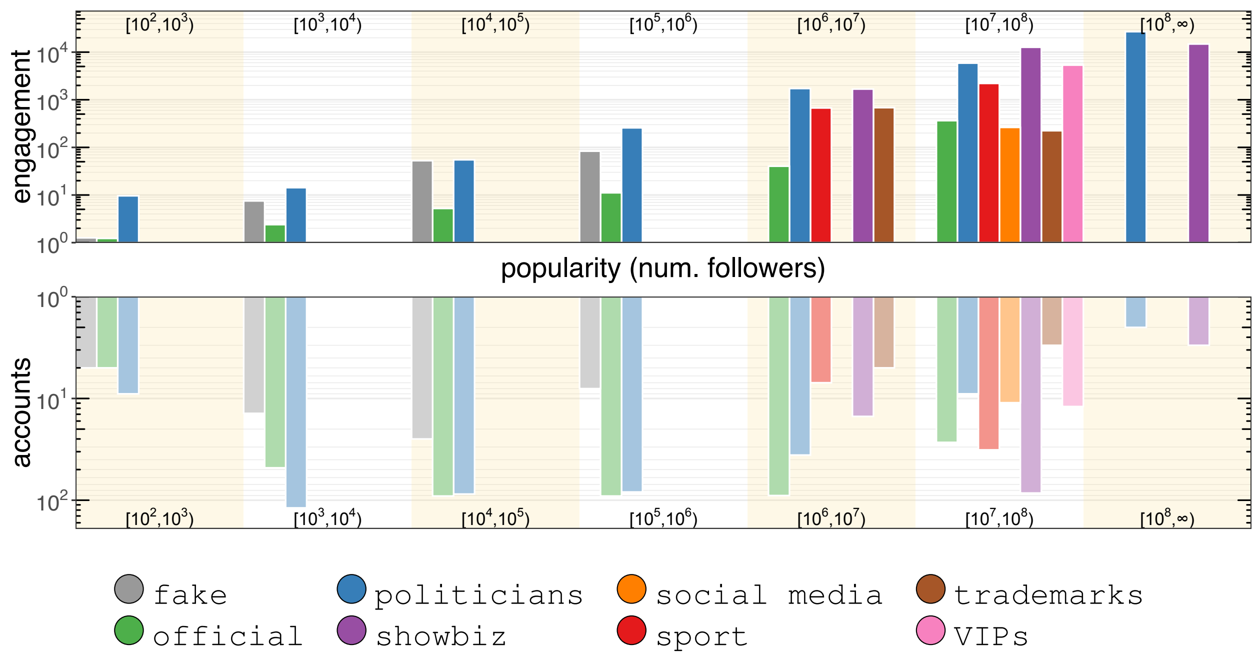


Fig. 3.8: Engagement obtained at different popularity levels by the different classes of actors. Although disinformation outlets (labeled `fake`) do not reach high popularity levels, they consistently obtain more engagement than official news outlets at middle and low popularity levels, and comparable engagement with respect to politicians.

CHAPTER 4

MEASURING THE EVOLUTION OF POLARIZATION AND NEWS INFLUENCERS BETWEEN TWO U.S. PRESIDENTIAL ELECTIONS ON TWITTER

In the previous chapter, we analyzed the impact of fake news during 2019 European election and showed that they had a marginal role in the online debate. This finding suggests that only one part of users endorsing a specific narrative interact with misinformation and fake news spreaders, while the majority of influential public personalities do not involve with such contents. Nevertheless, in some cases part of the research suggested that automated accounts may have strongly influenced the public debate by injecting misinformation, such as during the 2016 U.S. presidential election. During the last years, platforms increased the effort to limit the spread of misinformation and adopted several countermeasures such as the suspension of suspicious accounts and the reduction of the visibility of questionable contents. Thus, one question of interest is how the online debate has been affected from these measures over the last years.

In this chapter, we analyze the public debate around 2020 U.S. presidential election on Twitter and compare the results with the 2016 U.S. presidential election. We quantify the change in polarization between the two elections and provide insights about the evolution of the Twitter environment.

4.1 Introduction

A growing number of studies have documented increasing political polarization in the U.S. that is deeper than at any time since the American Civil War [79]–[81]. Partisan division over issues has increased among those affiliated with political and news media organizations – elected representatives, party officials, and political pundits – alongside an alarming increase in affective polarization among voters. This two-level pattern – issue polarization among political elites and affective polarization among voters – invites further research on the diffusion of polarized political information between those in positions of political influence and the larger population.

This diffusion of political information is difficult to track with traditional survey and roll call data that lack relational measures. Increasing reliance on social media for political communication is opening up unprecedented opportunities to study the diffusion of political information and misinformation [14] over communication networks [82]. Our study leverages social media data from Twitter to better understand the diffusion dynamics of news media information during the two most recent U.S. election cycles.

Twitter users are embedded in relatively stable communication networks created by the exchange of "retweets." A 2015 study by Metaxas et al. [83] found that "retweeting indicates not only interest in a message, but also trust in the message and the originator, and agreement with the message contents." The content of retweets makes it possible to identify information that is highly biased, as well as the ideological direction of the bias. Using retweet data we also can identify "influencers" who are the users with the greatest ability to broadly propagate new information over the retweet network. Typically, influencer tweets are highly likely to be retweeted, not only by their followers, but also by their followers' followers, and so on. We classify Twitter influencers into two categories. The first includes the "affiliated" who are associated with media or political organizations, and the other consists of the "unaffiliated" who do not have such associations, so most likely represent themselves or informal groups.

Our study also aims to better understand how polarization unfolds on social media. To clarify, political scientists distinguish multiple types and levels of polarization [84]–[89]: policy polarization (extreme differences of opinion on highly salient issues), partisan polarization (alignment of opinions with opposing political parties), ideological polarization (alignment of opinions with liberal vs. conservative world views), and geographic polarization (regional alignment of opinions, e.g., "red state/blue state"). Each of these four types of polarization can in turn be classified by level: elite polarization among political officials and pundits, media polarization among news organizations, and polarization among the underlying population as usually measured by exit polls and opinion surveys. In this paper, we use data from social media to study ideological polarization among the political elite, news organizations, and Twitter users more broadly. Over the past decade, the rapid growth of Twitter, Facebook, Reddit and other social media have transformed the communications and information propagation landscape. Alongside traditional broadcast media and face-to-face communication, people now have the ability to search for and exchange information with billions of other users in a global network. Recent studies have examined the impact of new technologies like Twitter and YouTube on election outcomes [90]–[99], including the effects of disinformation [16], [18], [19], [39], [100], [101]. Other studies have documented how social media platforms contribute to polarization through the creation of echo chambers [12], [102]–[110]. In contrast, here we focus on the increased polarization and involvement of Twitter influencers from from the 2016 to 2020 U.S. presidential elections. We measure longevity of Twitter influencers and the landscape of polarization of themselves and their retweeters during this period.

Our study focuses on the diffusion of news media information between influencers and those whom they influence, as well as the change in composition and popularity of these influencers and their retweeters. To maintain the consistency between the results from 2016 and 2020 elections, we follow the methodology of Ref. [18] to identify and classify the influences and their retweeters in the 2020 U.S. election data. We classify tweets containing a link to a news outlet into several news media categories corresponding to their different political orientations. We observe that the volume of tweets and users with a center orientation decreased from the 2016 election to the 2020 election. For each news media category, we reconstruct the corresponding retweet network and identify the most important news media influencers of the category by finding the most important nodes in terms of their ability to spread information in the network. The top 25 influencers in each news media category are then classified as affiliated with a media or with a political organization or unaffiliated. Finally, we measure the strength of the polarization of the influencers and of their retweeters, defined as the level of separation of the influencers' retweeters in two opposite clusters and find a clear, significant increase of the polarization from 2016 to 2020.

4.2 Results

4.2.1 News Media on Twitter in 2016 and 2020

We tracked the spread of political news on Twitter in 2016 and 2020 by analyzing two datasets containing tweets posted between June 1st and election day (November 8th in 2016 and November 2nd in 2020). The data were collected continuously using the Twitter Search API with the names of the two presidential candidates in each of the presidential elections in 2016 and 2020 as keywords. (Had we used more keywords targeting specific media outlets or hashtags concerning specific news events we would risk missing election-related tweets that did not contain references to the list of outlets or events.) The 2016 dataset contains 171 million tweets sent by 11 million users and was used in Refs. [18], [94] to assess the influence of disinformation on Twitter in 2016. The 2020 dataset contains 702 million tweets sent by 20 million users. Hence, we observe a significant increase in Twitter involvement in distributing election polarization, since in four years the number of Twitter users nearly double and number of tweets per user more than double, increasing the total number of tweets more than fourfold.

The classifications of news media websites presented below and used in this paper, including “fake”, “extremely biased”, “left”, “right”, and especially the boundaries between categories, are a matter of opinion, rather than a statement of fact. The categorizations and labels assigned to the corresponding classes used here originated in publicly available datasets from fact-checking and bias rating organizations credited below. The political views and conclusions contained in this article should not be interpreted as representing those of the authors or their funders.

For each tweet containing a URL link, we extracted the domain name of the

URL (e.g. `www.cnn.com`) and classified each link directing to a news media outlet according to this outlet's political bias. The 2016 and 2020 classifications rely on the website `allsides.com` (AS), followed by the bias classification from `mediabiasfactcheck.com` (MBFC) for outlets not present in AS (both taken as of January 7 2021 for the 2020 classification). We classified URL links in five news media categories for outlets that mostly conform to professional standards of fact-based journalism: *right*, *right-leaning*, *center*, *left-leaning* and *left*. We also include three additional news media categories to include outlets that tend to disseminate disinformation: *extreme-right bias*, *extreme-left bias* and *fake news*. Websites in the *fake news* category have been flagged by fact-checking organizations as spreading fabricated news or conspiracy theories, while websites in the extremely biased category have been flagged for reporting controversial information that distorts facts and may rely on propaganda, decontextualized information, or opinions misrepresented as facts. A detailed explanation of the methodologies used by AS and MBFC for rating news outlets and of the differences in classification between 2016 and 2020 is given in the Methods section. The full lists of outlets in each category in 2016 and 2020 are given in Tabs. A.1 and A.2. In the 2016 dataset, 30.7 million tweets, sent by 2.3 million users, contain a URL directed to a media outlet website. The 2020 dataset contained 72.7 million tweets with news links sent by 3.7 million users. This number reveals remarkable drop of fraction of flow of tweets from users associated with media form 18% in 2016 to 10% so nearly half lower. This came from mainly from smaller growth of productivity of media affiliated users.

The fractions of tweets and users who sent a tweet in each of the news media categories are shown in Fig. 4.1 A and B (the numbers are reported in Tab. A.3) along with other statistics about the activity of users in each category. Between 2016 and 2020, these fractions decreased most in the center category and increased most in the left-leaning category, with a smaller increase in the fractions in the right-leaning category. The shift away from the center may indicate the increasing polarization, both among users as well as media outlets. However, most of the decrease in the fraction of center media outlets reflects the shift of `CNN.com`, which was categorized by AS as center in 2016 and as left-leaning in 2020, combined with *CNN* accumulating more than twice the number of tweets in 2020 than the top outlet of the center category (`thehill.com`) (see Tab. A.2).

The fraction of tweets in the fake and extremely biased category, representing outlets that were most susceptible to sharing disinformation, decreased from 10% to 6% for fake news and from 13% to 6% for extremely right-bias news. The number of users who shared those tweets also decreased for extremely right-biased news (from 6% to 3%) but not for fake news (which remained at 3%) (see Tab. A.3). The fraction of tweets in the extremely-left bias category is very small (2%

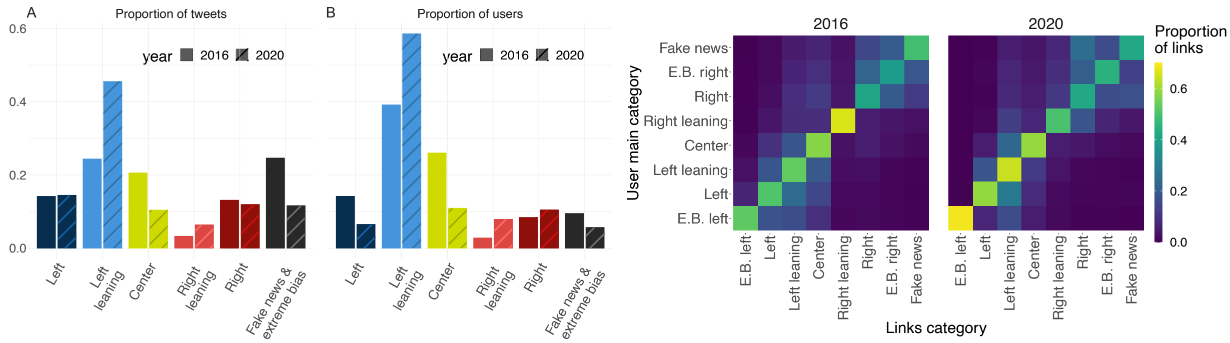


Fig. 4.1: **Distribution of news media links in 2016 and 2020, by news media category.** Panels A and B show the fractions of tweets and users that sent tweets with a URL pointing to a website belonging to one of the categories. Users are classified in the category in which they posted the most links. For the users that have at least two links classified, panels C and D report the fraction of links across categories as a function of the users' main categories.

in 2016 and even less, 0.05% in 2020).

Fig. 4.1 C and D shows the fraction of URLs across categories as a function of a user's modal category for users that posted at least two links in our datasets. The analysis reveals two clusters in 2016 and 2020, one with categories from the right and fake news (fake news, extreme-right bias news, right news and right-leaning news) and the second one with categories from the center and left (center news, left-leaning news, left news, extreme-left bias news). These two clusters can be interpreted as two echo chambers. Asymmetrical patterns in Fig. 4.1 C and D reveal that users in the right wing echo chamber also link to a very limited number of left wing media outlets, but that the opposite relation does not occur. This is consistent with asymmetry between left-leaning and right-leaning users in social media observed in previous studies [5], [18], [110].

In order to estimate the volume of tweets sent from automated accounts such as bots, we counted the number of tweets sent from unofficial Twitter clients, e.g., Twitter clients other than the Twitter Web client, Android client, IPhone client or other official clients. Unofficial Twitter clients include a variety of different applications used to automate all or part of an account activity, such as third party applications used typically by brands and professionals (e.g. SocialFlow or Hootsuite) or bots created with malicious intentions [94].

The overall fraction of tweets sent from unofficial clients was 8% in 2016 and 1% in 2020. A similar drop over the same period was observed in the average activity of these users (see Tab. A.3). This decrease could be attributed in part to measures taken by Twitter to limit the virality of disinformation. Our results show that while the relative volume of tweets linking to disinformation websites dropped approximately by a half in 2020 compared to 2016, the fraction of users sharing fake news decreased significantly (Fig. 4.1 A and B and Tab. A.3).

To understand how users shifted between categories from 2016 to 2020, we track users that are present in both election datasets and in both years are classified into the category in which they posted the most tweets in each year. Fig. 4.2 shows the resulting shifts. The two largest of them are of users that were in the center and left news category in 2016 and shifted to the left-leaning category in 2020. This reflects the consolidation of the left-leaning category as the largest in 2020, with the three most widely shared outlets: *New York Times*, *Washington Post* and *CNN* (see Tab. A.6). We also observe a large fraction of users in the fake and extremely biased news category in 2016 that moved to the right news category in 2020. However, these user shifts also reflect the change in the classification of media outlets from 2016 to 2020. We infer the ideological position of Twitter users without relying on the news outlet classification in section 4.2.3, and show that the resulting positions are highly correlated with the user positions computed using the news categories in which they posted.

4.2.2 News Media Influencers

To capture the dynamics of information diffusion, we reconstruct retweet networks corresponding to each news media category. We add a link, or directed edge, going from node v to node u in the news network when user u retweets the tweet of user v that contains a URL linking to a website belonging to one of the news media categories. With this convention, the direction of the link represents the direction of the influence between Twitter users. We do not include multiple links with the same direction between the same two users or self-links (when a user retweets their own tweets). The in-degree of a node is the number of links that point inward to the node and its out-degree is the number of links that originate at a node and point outward to other nodes. With our convention, the in-degree of a user is equal to the

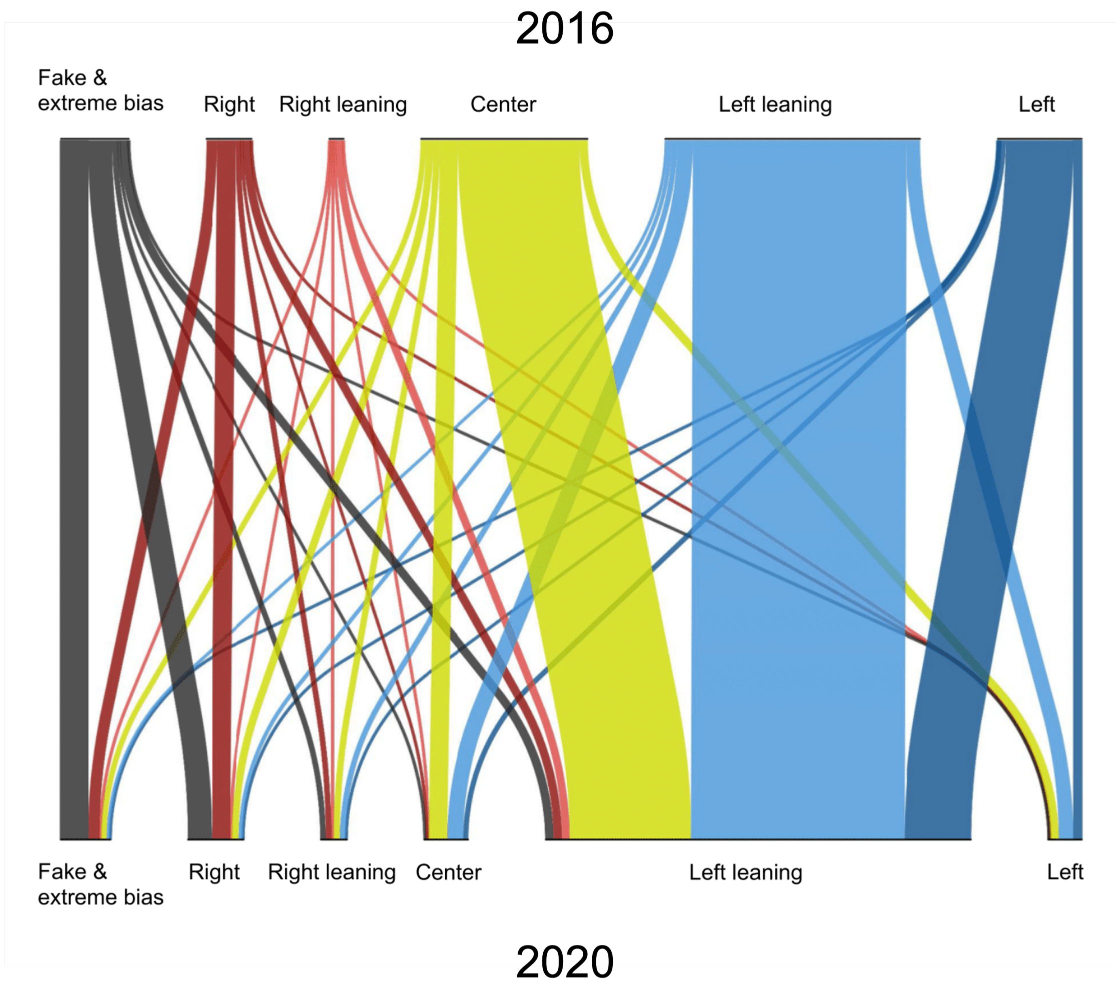


Fig. 4.2: **Shifts of users across news media categories from 2016 to 2020.** The size of each category in 2016 corresponds to the number of unique users in the category in 2016 (Fig. 4.1). The shift from one category to another is proportional to the fraction of users that were classified in 2016 and in 2020 in the two respective categories.

number of users they retweeted at least once and their out-degree is the number of users who have retweeted them at least once. The higher a node's out-degree, the greater its local influence. The characteristics of the retweet networks are given in Tab. A.4.

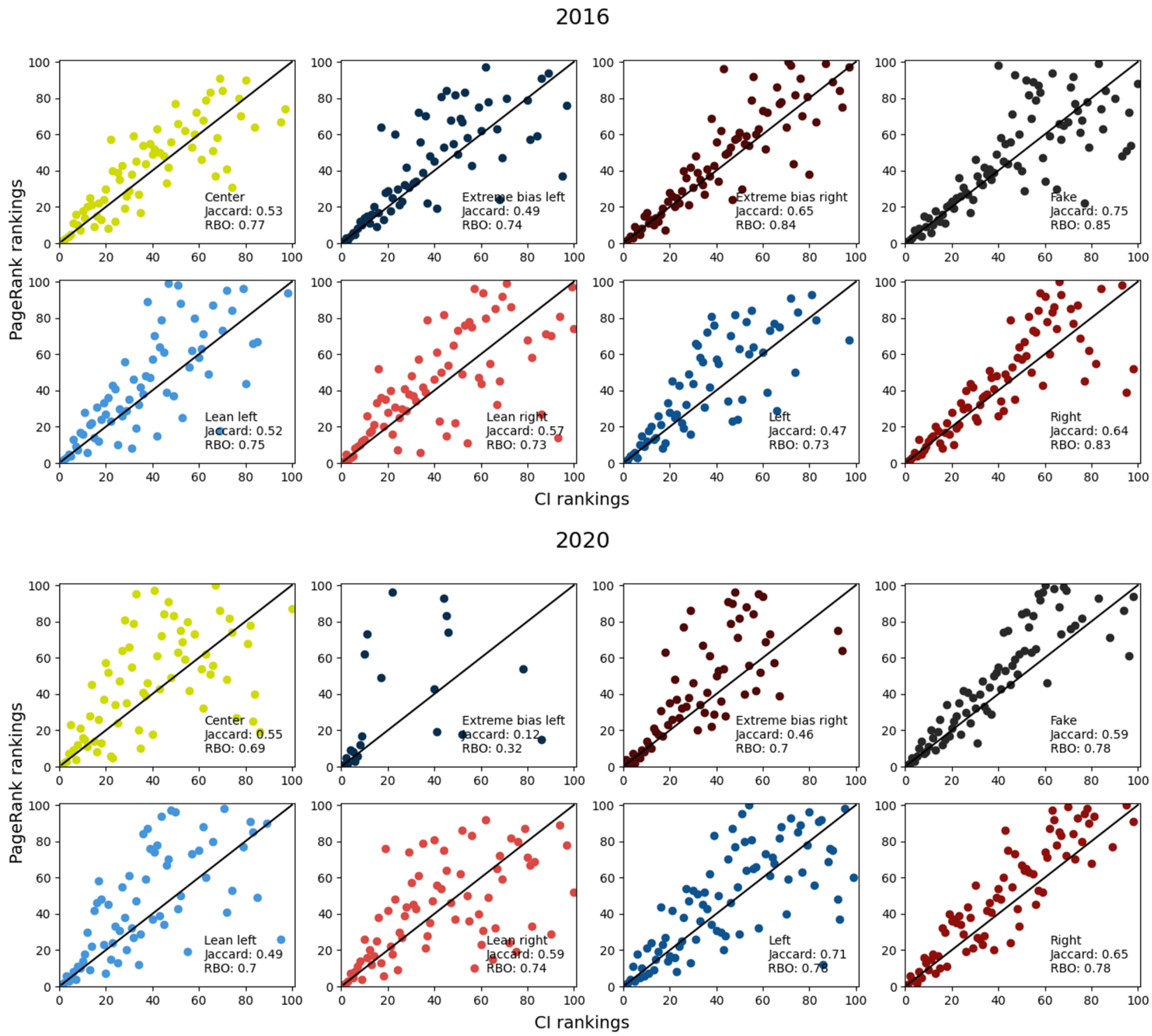


Fig. 4.3: **Comparison of top 100 rankings generated by the PageRank algorithm and by the Collective Influence (CI) algorithm using the 2016 and 2020 retweet networks.** Ranked Bias Overlap (RBO) [111] and Jaccard Similarity are computed over the two top 100 lists, shown below their respective news category labels. For this analysis, RBO's weight parameter p is set to 0.98. The RBO values are generally above 0.7 indicating a high agreement of the two ranking, especially for the top ranked users. The only network that show a poor agreement between the rankings is the extreme bias left network of 2020. This may be explained by the small size and low average degree of the network compared to networks of other categories (see Tab. A.4).

In each retweet network, we use the Collective Influence (CI) algorithm [112] to find the best spreaders of information, i.e. the *influencers* of the corresponding news media category. Specifically, the CI algorithm finds the minimal set of nodes that can collectively reach the entire network when information diffuses according to a linear threshold model. The CI algorithm considers influence as an emergent collective property, not as a local property such as the node's out-degree. It does this by finding the smallest set of nodes needed for global cascades. Accordingly, the CI algorithm is able to rank super-spreaders of information in social networks [18], [113], [114].

Here, we use a directed version of the algorithm to identify the super-spreaders of information as the nodes with the highest CI_{out} to be able to compare results from both elections [18]. The 2016 influencers' rankings are shown in the upper panel in Fig. 4.4 for the top five influencers, and in Tab. A.5 for the top 25 influencers. Analysis of these results reveals that traditional news influencers were mostly journalists with verified Twitter accounts linked to traditional news media outlets. In contrast, fake and extremely biased news are sent mainly by influencers whose accounts are unverified or deleted, with deceptive profiles and much shorter life-span on Twitter than traditional media influencers. However, some of these influencers, despite their unknown, non-public nature, still played a major role in the diffusion of disinformation and information on Twitter [18].

The results from analysis of 2020 data are shown in the bottom panel of Fig. 4.4. For influencers that persisted from 2016, their previous position in 2016 is listed in purple parentheses (see also Tab. A.6). Those influencers are often highly ranked in both the 2016 and the 2020 analyses. Among the union of top 100 influencers from each category in 2020 (representing 598 unique users) 150 were already in the top 100 of one category in 2016. Yet, this means that 75% of the top 100 influencers in 2020 are new to such high ranking.

The CI algorithm operates on the unweighted retweet networks. To verify that a ranking computed on the weighted networks would not produce significantly different results, we compare the CI ranking with the ranking obtained from the PageRank algorithm applied to the weighted networks. The comparison reveals a strong agreement, especially for highly-ranked users as shown in Fig. 4.3

Fig. 4.4 shows the retweet networks for each news media category in 2016 and 2020, among communities formed by the union of the top 30 influencers for each category. The two force directed network layouts are computed using the same parameters and show two main clusters, with the right-biased and fake news influencers in one cluster and the left-biased influencers in the other. The increased separation in 2020 is notable. In 2016 the center influencers are mostly between the two clusters; in 2020 the separation between the two clusters increased and only a few influencers remain within a central position (e.g. `@thehill`). We quantify the polarization of the full set of top 100 influencers and of their retweeters, using all the retweets between them, in detail in the next section.

Fake and extremely biased news are sent mostly by influencers whose accounts are unverified or deleted, with fake news seeing a significant increase in deleted influencer accounts from two in the top 25 in 2016 to eight in 2020 (see Tabs. A.5 and A.6). Conversely, the extreme right-biased news in 2020 consisted primarily of verified influencers that grew from 15 verified in the top 25 in 2016 to 23 in 2020.

Using a manual labeling process (see section 4.4), we classify the top 25 influencers of each news media category for 2016 and 2020 as affiliated with media or political organizations, or unaffiliated, in order to observe the makeup of influencer types for these categories. An influencer affiliated with a media organization could be a media company or official media outlet (e.g. `@FoxNews`), or a writer, reporter, consultant, or other individual who has directly corresponded with a media outlet in an officially recognized capacity (e.g. `@joelpollak`). An influencer affiliated with a political party could be a politician (e.g. `@JoeBiden`), a political campaign platform or an affiliate of the platform, or someone who officially represents an aspect of U.S. politics (e.g. `@joncoopertweets`). We also split the unaffiliated category into two subcategories: independent and "other." An independent influencer is an influencer not officially affiliated with any media or

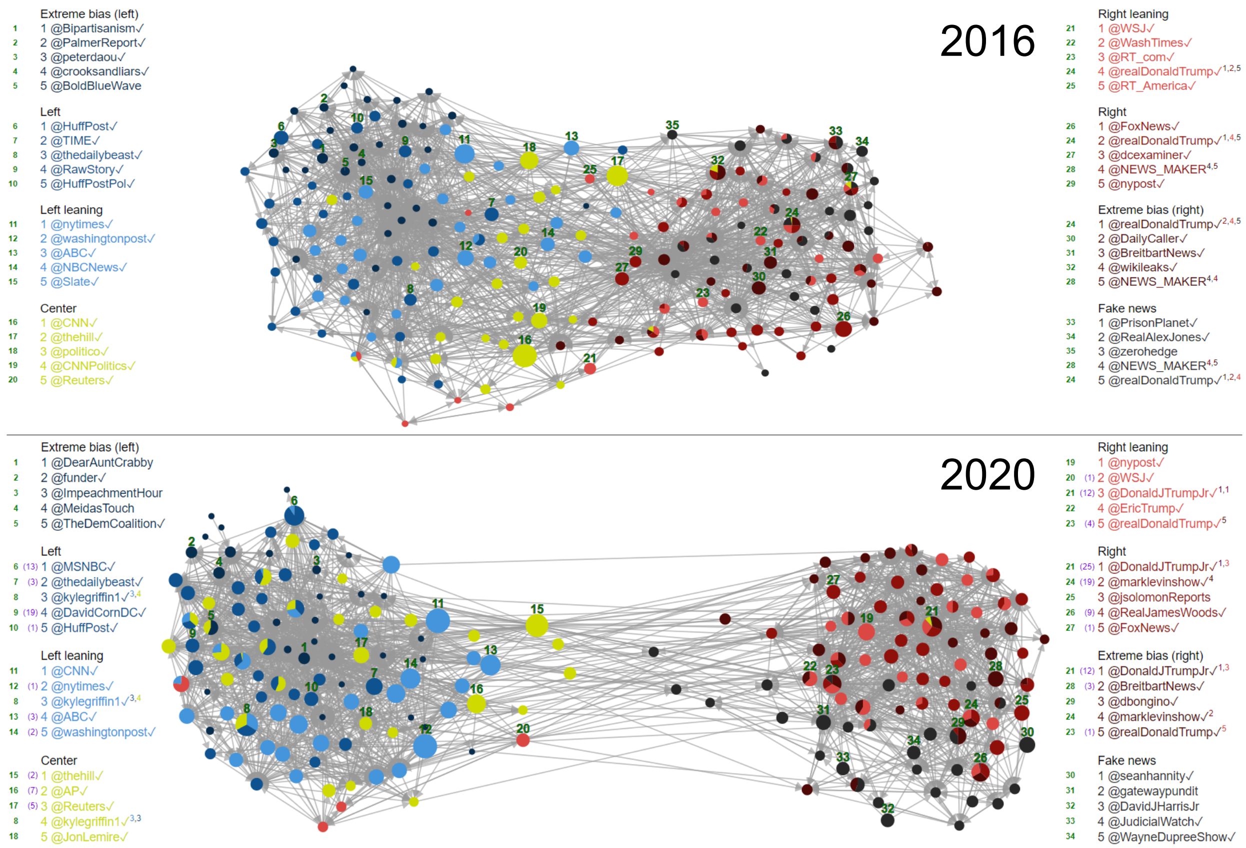


Fig. 4.4: **Retweet networks formed by the top 30 influencers within each media category, by year.** The 2016 network (upper panel) was generated from 2016 data using the same algorithm as used in [18] but with different parameters to ease its comparison with the bottom panel generated from 2020 data. The arrows show the directions of links between users, from the source of influence (the original poster) to the recipient (the retweeter). The size of a node is proportional to its out-degree in the complete combined network, i.e., the number of different users that have retweeted the node at least once with a URL directing to a media outlet. The color of a node indicates the news media category with which the node is affiliated. Nodes ranked in the top 30 of multiple categories are represented by pie charts where the size of each slice is proportional to their CI_{out} ranking (i.e. the node's collective influence). Both networks are visualized using a force-directed graph layout. The tables on either side of the networks show the top five users in each news media category. The number in green to the left of each user is their unique index, used to label the user's node in the network. Users ranked in the top 30 for multiple news media categories have colored superscripts, indicating the rank and media classification of their other top five positions. Verified users are indicated by a checkmark ✓. In the 2020 tables, a user's 2016 rank is displayed with the purple number to the left of their 2020 rank. Three usernames in the top 10 changed between 2016 and 2020: `@DRUDGE_REPORT` became `@NEWS_MAKER`, `@HuffingtonPost` became `@HuffPost`, and `@TruthFeedNews` became `@TAftermath2020`. Those users will have their new handle displayed in the 2016 tables for consistency (as well as in Figs. 4.7 and 4.9).

political platforms (e.g. `@amberofmanyhats`). The "other" category represents influencers whose accounts have no descriptions or context that could be used to identify them.

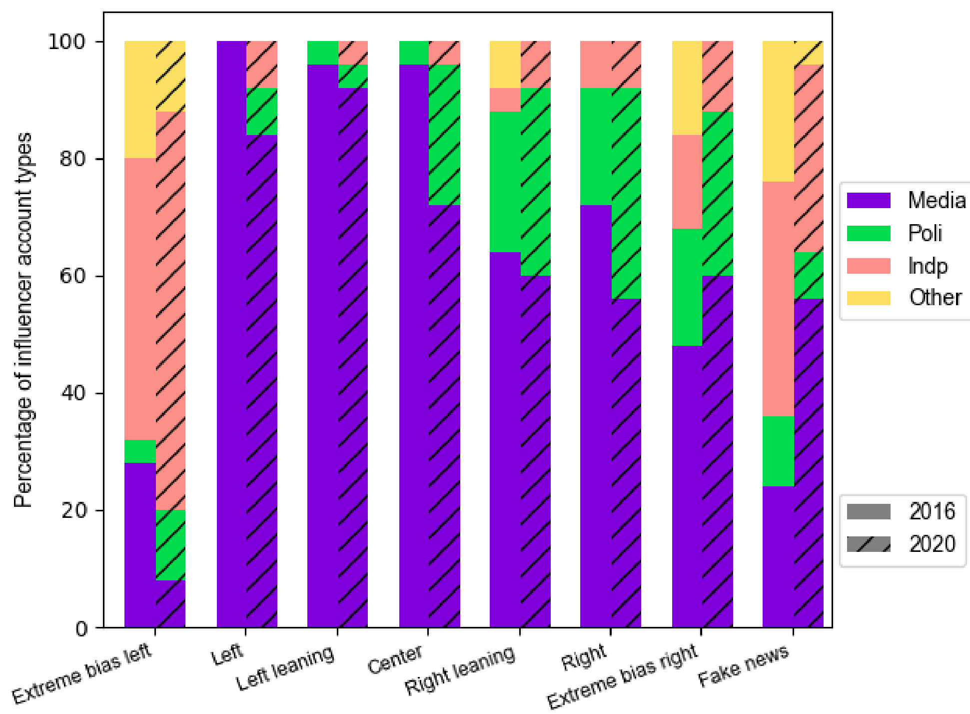


Fig. 4.5: **Reshuffling of distribution of the top 25 influencer types from 2016 to 2020, by news media category.** Influencers are classified as affiliated with a media organization, political organization, independent, or other (e.g. unidentified).

The fractions of influencers in these categories are shown in Fig. 4.5. It reveals that unaffiliated influencers are more common in the fake and extreme-bias categories, while affiliated influencers are more common in the other news categories. A similar trend is evident in the fractions of verified and unverified influencers found in these categories (see Tab. A.6), as fake and extreme-bias news categories

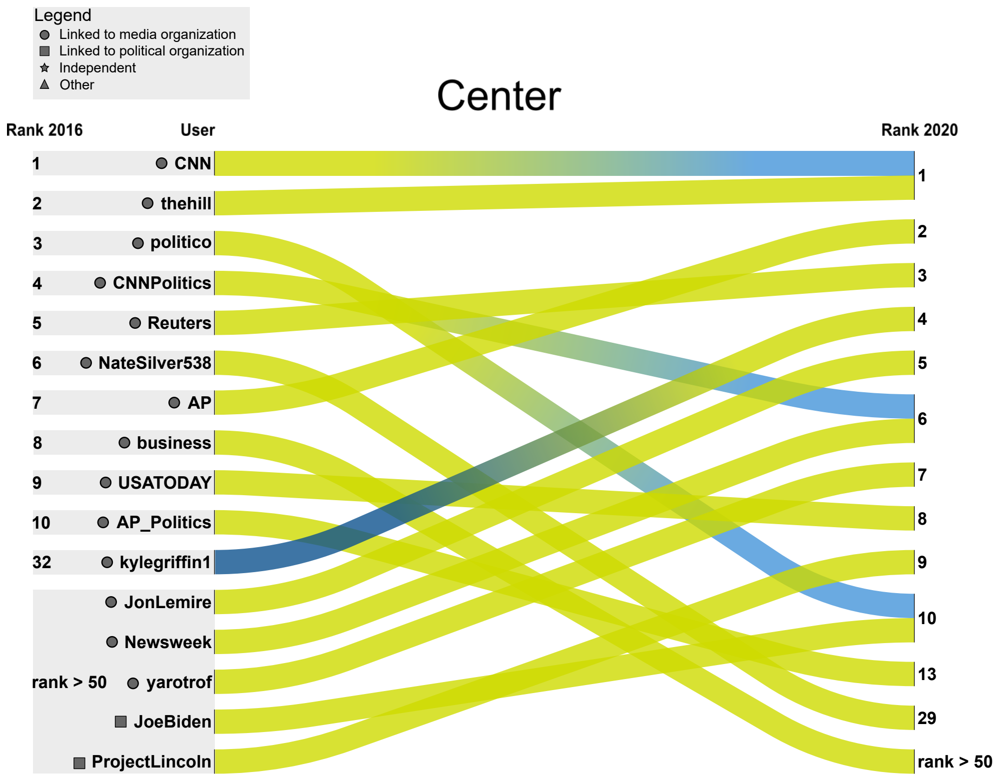


Fig. 4.6: **Change in rankings 2016-2020, Center Bias**. Outlines the change in the ranks of the top 10 center bias users from 2016 and 2020, ranked by CI influence. Each flow connects the best ranking for a user in 2016, whose rank is displayed to the left of the user handle, to their rank in 2020. The color of the lines match the bias of the users best ranking, and gradients represent a change in the bias classification of their best ranking. Note user @kylegriffin1 is more highly ranked in the left leaning bias (rank 3) but we chose to show its center ranking for this center bias plot, as the difference in rank is small and it keeps the figure focused on the center bias.

generally contain fewer verified influencers. In addition, media-affiliated influencers seem to have a greater presence in the left, left-leaning, and center news categories compared to their counterparts. Interestingly, the number of media-affiliated influencers within most of the categories actually decreases from 2016 to 2020. The exceptions are the extreme-right bias and fake news categories, which actually increased in media-affiliated influencers, while the extreme-right bias category also increased in politically-linked influencers. This indicates a shift in polarization of influencers affiliated with right-biased political and media organizations toward the extreme-right bias and fake news, as well as the emergence of new media-affiliated influencers in these categories. We discuss these changes in more detail below. In addition to changes in user types and verified users from 2016 to 2020, we observe a significant reshuffle of the ranking of influencers. Fig. 4.7 shows the change in rankings of the top 10 influencers in left and left-leaning, right and right-leaning and extreme-right bias and fake news categories. The ranking reshuffle in the center news category is shown in Fig. 4.6.

The comparison reveals several interesting changes between 2016 and 2020. First, we see that highly influential users rise from obscurity. Across all categories, a set of previously unranked or very low-ranked users break in to the top-10 rankings. These users include, for example, `@TeaPainUSA`, `@svdate`, `@kylegriffin1`, `@marklevinshow`, `@DavidHarrisJr`, etc. Considering all unique users in the top 25 influential users (across all categories of news media), we see that 58% came from outside the top 100 influential users in 2016. However, most of these newly influential users are related in some way to media or political organizations, while 28% of these new influencers are independent.

Observing the change in rankings by news media category, we see that right/right-leaning and extreme-right bias/fake news categories have a significantly higher fraction of top 10 influencers who were previously outside the top 50, compared to the change in rankings among the groups in left/left-leaning news categories. All categories show a large number of influencers falling out of the top 50 from 2016 to 2020, and in the case of the left news influencers, we see their former positions filled by users who were much less influential in 2016. The influencers with extreme right bias and fake news affiliations show the most volatility with regards to retaining top-10 influencer positions, with many top-10 influencers in 2016 ranked below 50 in 2020 (or were even banned from Twitter, like `@RealAlexJones`).

The change of classification of some news media outlets is also reflected in the category shifts of their Twitter accounts. In particular, `@CNN` and `@politico` – previously the first and third highest ranked influencers in the center category in 2016 – shifted to left-leaning. Such shifts of large and influential media influencers across news categories indicates the increased content polarization on Twitter. A shift of media-affiliated influencers from the right to the extreme right is also visible (e.g. `@DailyMail`, `@JudicialWatch`, `@marklevinshow`), as well as the emergence of new-media affiliated influencers in these categories (e.g. `@newsmax`, `@OANN` or `@RaheemKassam`). In contrast to the shift to the extremes among large media influencers, the center rankings remained fairly consistent between 2016 and 2020 as shown in Fig. 4.6. Some new users rose from low ranks to fill in the gaps, including `@JoeBiden`, but only one user dropped

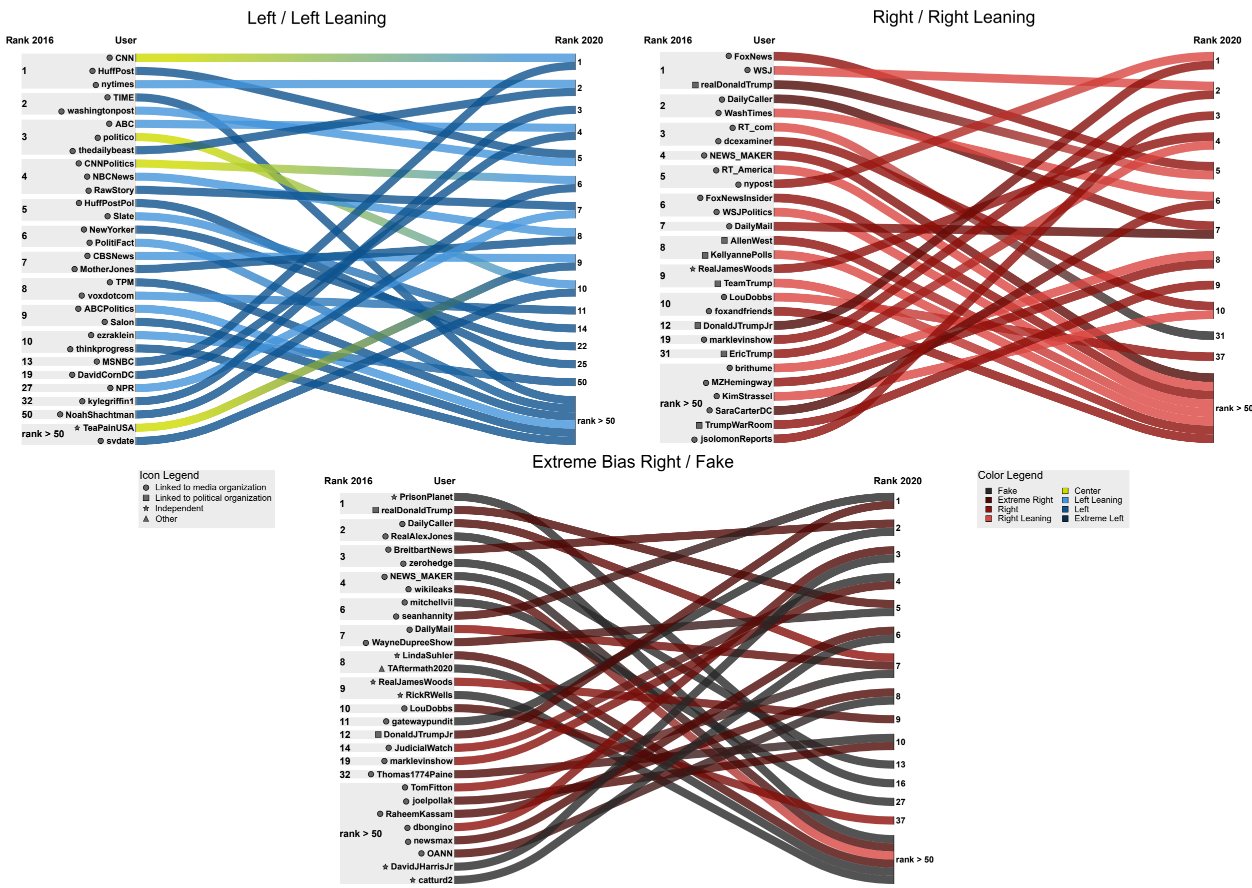


Fig. 4.7: **Change in influencers rankings from 2016 to 2020.** Influencers ranked in the top 10 in at least one news media category in 2016 or 2020 are shown. The 2016 rankings are displayed to the left of the username, with 2020 rankings on the right. For each user only one shift is shown. Its color changes from the user's highest ranked news media category in 2016 to such color for 2020. Each panel shows the change over time between two news media categories.

out of the top 50 entirely, and the remaining shifts are internal to these top-ranked users.

4.2.3 Polarization among Twitter Users

The evolution of influencers across different news media categories (Figs. 4.1 and 4.2) suggests an increased polarization in the relations among influencers between

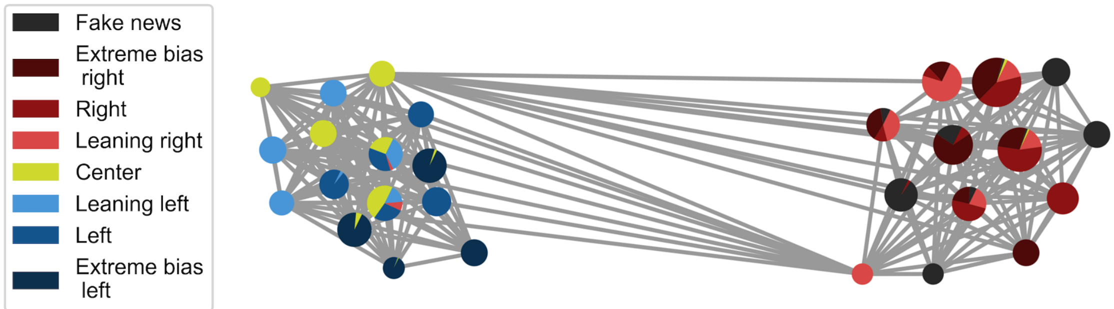


Fig. 4.8: **Similarity network for a random subsample of the 2020 influencers**. Each edge is weighted by the cosine similarity between retweeting users. Size of the node represents that node's degree centrality. The pie charts representing the nodes illustrate the news categories to which that node belongs, with the size of the slices denoting their relative influence for that category. For clarity, edges below the average inter-community edge weight are hidden. Nodes are grouped relative to each other by their detected community.

2016 and 2020. Here we broaden the scope of polarization analysis to the Twitter users who are consuming and retweeting the influencers' content. For the 2016 and 2020 data, we consider the union of the top 100 influencers of each news media category as a single set representing the most influential users covering the entire news media ideological spectrum for the target year. For this analysis we use all the retweets in our datasets, not only those containing a link to a news outlet, and remove the ones sent from unofficial Twitter clients to capture only tweets sent by humans. Using these influencers as nodes, we create two fully connected similarity networks derived from the 2016 and 2020 Twitter network, respectively. An edge between any two influencers in the created networks represents similarity of the number of retweets of these two influencers for every user in the corresponding Twitter network (see Methods for more details). In both similarity networks, a community detection algorithm found two communities. One contained influencers affiliated with news media in the center, left-leaning and left news categories, while the other contained those affiliated with news media in the right-leaning, right and fake news categories. This indicates that influencers separate their user bases according to the content they generate. We illustrate this separation in Fig. 4.8 that shows a sample of the 2020 similarity network. To quantify the difference in community separation and, subsequently, polarization, between the two networks, we measured the modularity and normalized cut between communities (see section 4.4 for details).

The modularity for the 2020 network was 0.39 with a standard error (SE) of 0.01, versus 0.365 (SE = 0.007) in 2016, indicating more closely knit communities in 2020, with stronger in-community ties and weaker between-community ties. Consistent with the increase of community modularity, the average normalized cut of 0.36 (SE = 0.04) in 2016 decreased to 0.128 (SE = 0.005) in 2020. To interpret this change, we note that on average, each node in the 2016 similarity network had 64% of in-community edges and 36% of across-community edges. The latter fraction decreased to 13% in 2020, dropping nearly three times lower than it was in 2016. This indicates much stronger separation of these communities in the later election. We also computed the above metrics on networks generated from user quote similarity in order to show that retweets are the strongest form of endorsement of influencer content, and subsequently the best approach for our analysis (see Tab. A.7).

To quantify and compare the polarization not only among Twitter influencers but also among the users, we infer the ideology of Twitter users based on the ideological alignment of political actors they follow [12], [47]. The bipartite network of followers is then projected on a one dimensional scale using correspondence analysis [48], [49], which applies a SVD decomposition of the adjacency matrix standardized to account for the differences in popularity and activity of the influencers and their followers (see section 4.4 for details). Two users are close on the resulting latent ideology scale if they follow similar influencers. This method has been shown to produce ideological estimates of the members of the U.S. Congress highly correlated with ideological estimates based on roll call voting similarity such as DW-NOMINATE [47].

For 2016 and 2020, the data for the analysis consists of the union of the top 100 influencers of each news media category and the sets of users that retweeted at least three different influencers (considering all tweets in our datasets, not only the ones with URLs). Following the finding in [83] that "retweeting indicates not only interest in a message, but also trust in the message and the originator, and agreement with the message contents," we interpret retweeting a form of endorsement of the content being retweeted. Twitter offers other types of interactions allowing users to comment on the content, such as quote tweets and replies. The ratio of quotes to retweets of users to influencers was very stable and small ($< 5\%$) in 2016 and 2020, for users on the left and right sides of the latent ideology (see Tab. A.8-A), which motivated our focus on retweets to infer the ideology of users. We note that the ratio of quotes to retweets from users of one side of the ideology spectrum to influencers of the other side increased from 2016 to 2020, indicating an increased usage of quotes to comment on tweets from influencers of the opposite side. However, the overall usage of quotes over retweets remained small (see Tab. A.8-B). We extract the coordinates of each user on the first dimension of the results of the correspondence analysis applied to the weighted network of retweets between the users and the influencers (see section 4.4 for the details and robustness checks that we performed). Finally, for 2016 and 2020, the coordinates of all users are standardized to a mean of zero and a standard deviation of one. Two users are close together on the latent ideology scale if they tend to retweet similar influencers. The influencers' latent ideological positions are then computed as the median of their retweeters' positions.

Fig. 4.9 shows the result of this analysis. The latent ideology of the top five influencers of each category is shown as a box plot representing the distribution of the ideology of the users who retweeted them. The distribution of ideology positions of the users and of the influencers, displayed in green and purple, re-

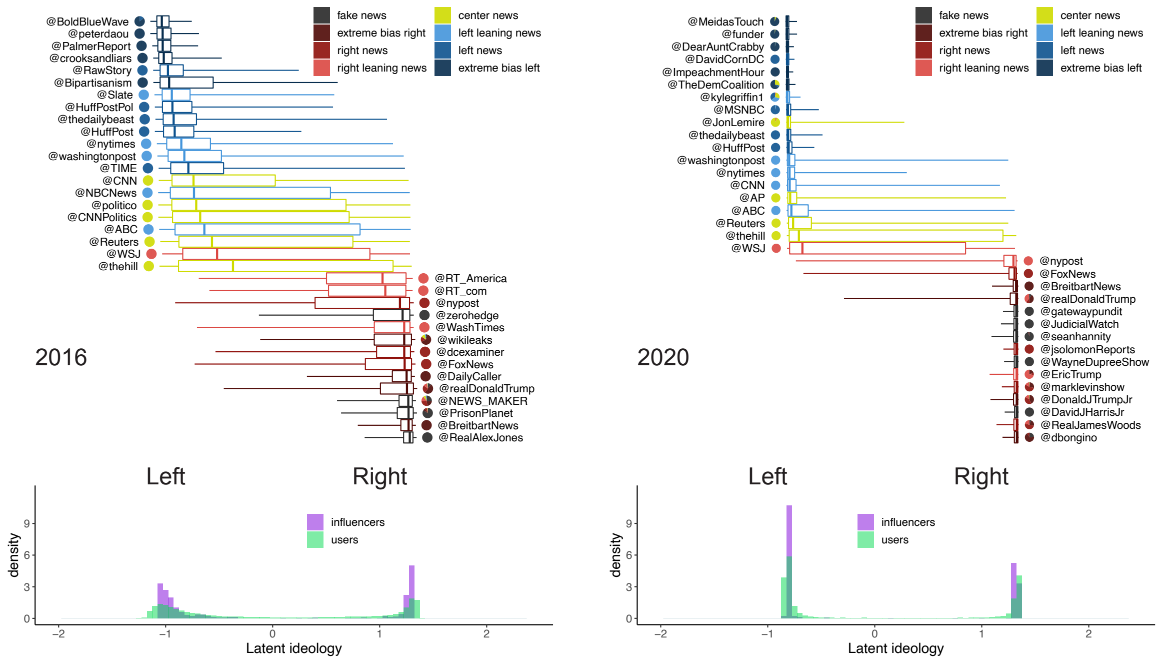


Fig. 4.9: **Latent ideology scale of influencers and their retweeters in 2016 (left) and 2020 (right)**. The latent ideology of the top five influencers of each category is shown as a box plot representing the distribution of the ideology of the users who retweeted them. The distributions for the users are shown in green, and the distributions for the top 100 influencers of each news media category (computed as the median of the ideology of their retweeters) are displayed in purple. Box plots indicate the 25% and 75% percentiles of the distributions with whiskers indicating the 5% and 95% percentiles. Pie charts next to the influencers' names represent the news categories to which they belong (weighted by their respective CI ranks in each category).

spectively, shows that polarization increased between 2016 and 2020. This is confirmed by a Hartigans' dip test (HDT) for unimodality, which measures multimodality in a sample by the maximum difference, over all sample points, between the empirical distribution function, and the unimodal distribution function that minimizes that maximum difference [115]. For the user distribution, the test statistics is $D = 0.1086$ (95% CI: [0.108,0.109]) in 2016 and $D = 0.1474$ (95% CI: [0.1471,0.1477]) in 2020. For the influencer distribution, the test statistics is

$D = 0.17$ (95% CI: [0.16,0.20]) in 2016 and $D = 0.21$ (95% CI: [0.19,0.23]) in 2020. All tests reject the null hypothesis of a unimodal distribution with $p < 2.2 \times 10^{-16}$ and the 95% confidence intervals are computed from 1000 bootstrap samples using the bias-corrected and accelerated method. Increasing values of the test statistic indicates distributions that increasingly deviate from a unimodal distribution, corroborating the growing division found in the similarity networks.

To understand if the measured increase in polarization is due to the arrival of new users and influencers in 2020, we repeat this analysis including only users (shown in Fig. 4.10), only influencers (Fig. 4.11) or only users and influencers (Fig. 4.12) that were active during both elections. In all cases we observe an increase of the Hartigans' dip test (HDT) statistics (see Fig. 4.13 and Tab. A.9) indicating that the increased polarization is not only due to the departure and arrival of new users between elections but also to a change of behavior of the users that remained. The largest increase in HDT for the user distribution is obtained when all users of 2016 and 2020 and only influencers that were present during both years are considered ($+0.08$). This setting also corresponds to the smallest increase of the dip test of the influencer distribution ($+0.01$, within 95% CI), suggesting that the new influencers of 2020 have more polarized ideologies than the influencers who remained from 2016 and that the increased polarization of the users is due in large part to the arrival and departure of users between elections (see Fig. 4.13 and Tab. A.9).

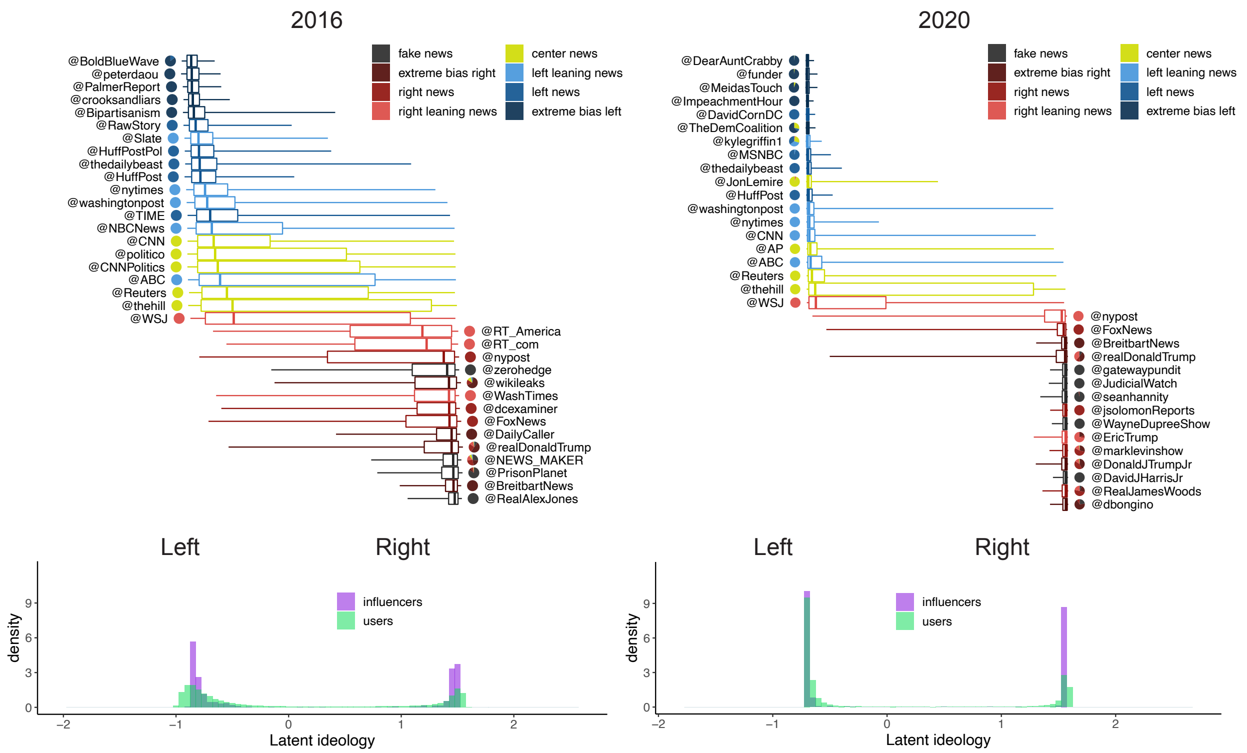


Fig. 4.10: **Latent ideology scale of influencers and their retweeters in 2016 (left) and 2020 (right) using only users active in both years**. The latent ideology of the top 5 influencers of each category is shown as a box plot representing the distribution of the ideology of the users having retweeted them. The distribution of the ideology estimate of the users is shown in green and the distribution of the ideology estimate of the top 100 influencers of each news category (computed as the median of the ideology of their retweeters) is displayed in purple. Pie charts next to the influencers' names represent the news categories they belong to (weighted by their respective CI ranks in each category). Hartigans' dip test for unimodality applied to the user distribution is $D = 0.094$ ($p < 2.2 \times 10^{-16}$) in 2016 and $D = 0.117$ ($p < 2.2 \times 10^{-16}$) in 2020. The test statistics for the influencer distribution is $D = 0.178$ ($p < 2.2 \times 10^{-16}$) in 2016 and $D = 0.214$ ($p < 2.2 \times 10^{-16}$) in 2020.

Figure 4.9 reveals a clear increase in polarization of the users and influencers in 2020 compared to 2016 and an alignment of their latent ideologies in two distinct groups, mirroring the news media classification groupings seen in Fig. 4.4 and Fig. 4.8. The box plots show that the distributions of users retweeting in-

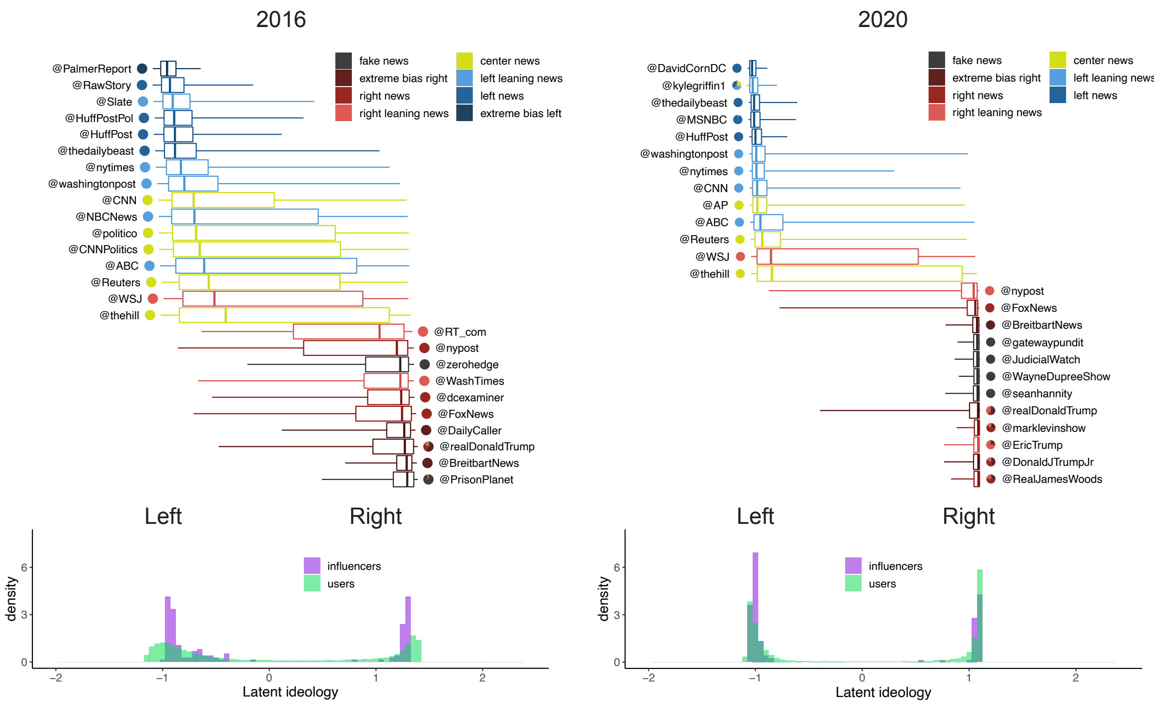


Fig. 4.11: **Latent ideology scale of influencers and their retweeters in 2016 (left) and 2020 (right) using only influencers active in both years**. The latent ideology of the top 5 influencers of each category is shown as a box plot representing the distribution of the ideology of the users having retweeted them. The distribution of the ideology estimate of the users is shown in green and the distribution of the ideology estimate of the top 100 influencers of each news category (computed as the median of the ideology of their retweeters) is displayed in purple. Pie charts next to the influencers' names represent the news categories they belong to (weighted by their respective CI ranks in each category). Hartigans' dip test for unimodality applied to the user distribution is $D = 0.107$ ($p < 2.2 \times 10^{-16}$) in 2016 and $D = 0.183$ ($p < 2.2 \times 10^{-16}$) in 2020. The test statistics for the influencer distribution is $D = 0.163$ ($p < 2.2 \times 10^{-16}$) in 2016 and $D = 0.173$ ($p < 2.2 \times 10^{-16}$) in 2020.

fluencers became more concentrated in 2020, with two clear opposite poles and fewer influencers having a user base bridging opposite ideologies. These results independently confirm the shift of news outlets and influencers from the center to the right and left observed using the news media classifications by external sources. Indeed, we find an extremely high correlation (above 0.90 for 2016 and

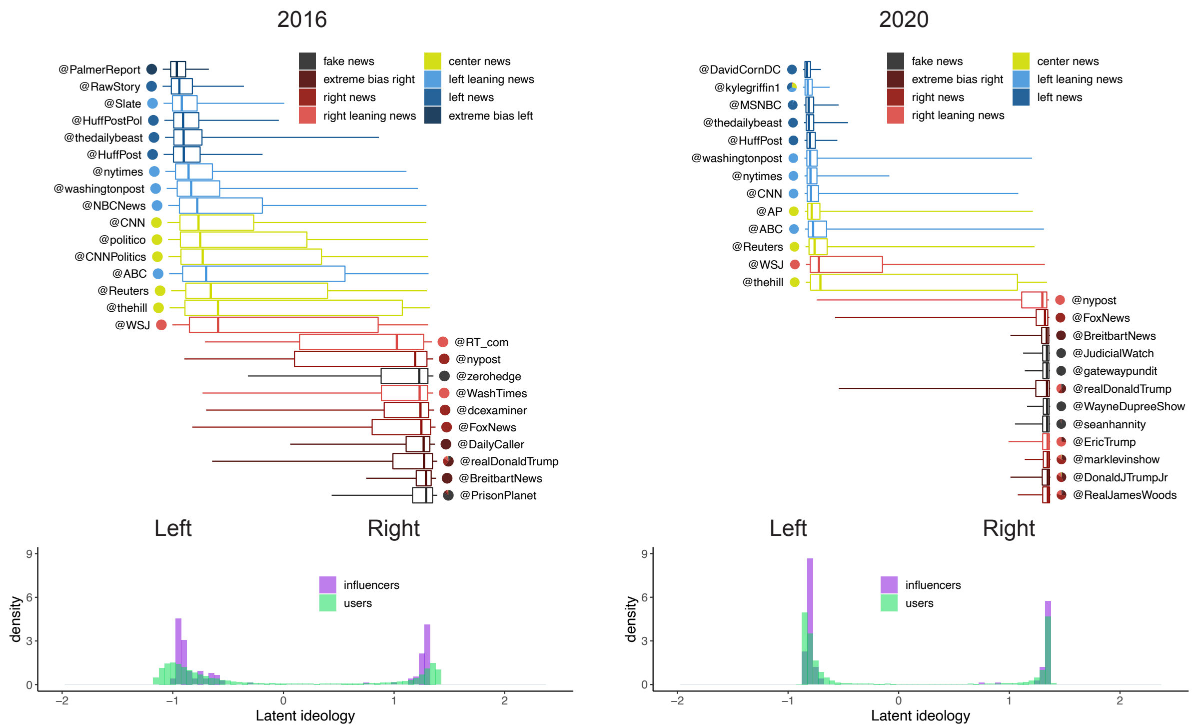


Fig. 4.12: **Latent ideology scale of influencers and their retweeters in 2016 (left) and 2020 (right) using only users and influencers active in both years**. The latent ideology of the top 5 influencers of each category is shown as a box plot representing the distribution of the ideology of the users having retweeted them. The distribution of the ideology estimate of the users is shown in green and the distribution of the ideology estimate of the top 100 influencers of each news category (computed as the median of the ideology of their retweeters) is displayed in purple. Pie charts next to the influencers' names represent the news categories they belong to (weighted by their respective CI ranks in each category). Hartigans' dip test for unimodality applied to the user distribution is $D = 0.095$ ($p < 2.2 \times 10^{-16}$) in 2016 and $D = 0.140$ ($p < 2.2 \times 10^{-16}$) in 2020. The test statistics for the influencer distribution is $D = 0.164$ ($p < 2.2 \times 10^{-16}$) in 2016 and $D = 0.171$ ($p < 2.2 \times 10^{-16}$) in 2020.

2020) between the users' latent ideology position and their left- or right-leaning distribution computed using the news media categories in which they posted (see section 4.4). This high correlation indicates that the shift in bias observed at the level of the media outlets is also present at the level of the users' retweeting pattern and serves as an independent validation of the media outlet classification.

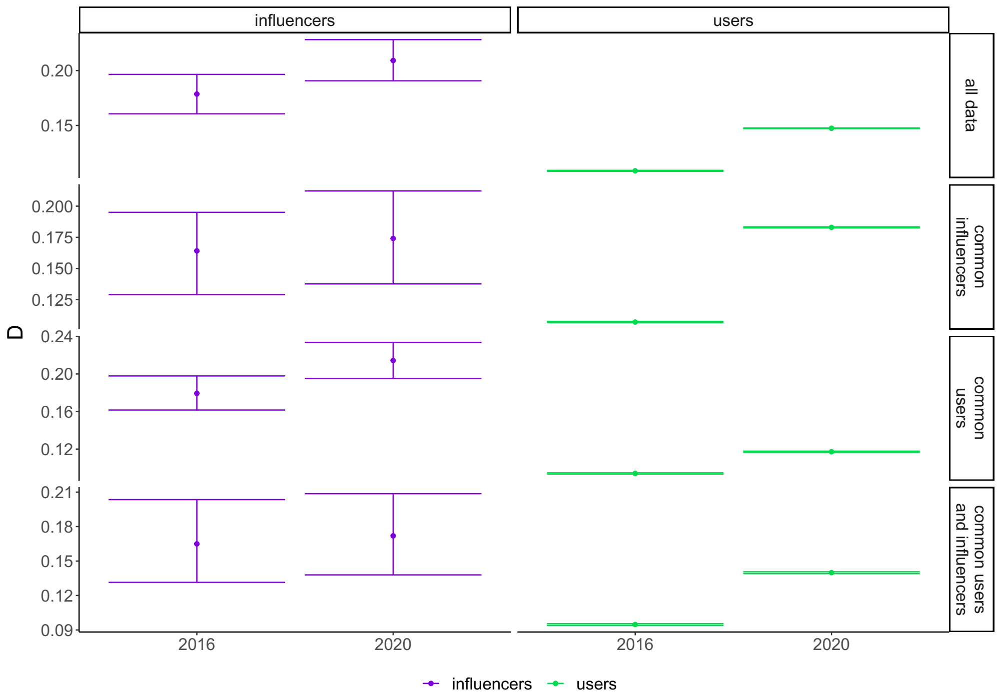


Fig. 4.13: Hartigans' dip test values for ideology distribution of users and influencers when considering all users and influencers or only influencers or users present in 2016 and 2020. 95% CI error bars are obtained by bootstrap with 1000 runs for each dataset and Bias-corrected and accelerated confidence intervals method. The numerical values are reported in Table A.9.

4.3 Discussion

This work uses Twitter retweets to study polarization among influencers and those they influence in the months leading up the 2016 and 2020 U.S. Presidential elections. Multiple analyses confirm a robust pattern of increasing division into opposing echo chambers, largely due to the arrival of new, more polarized influencers and users in 2020. Among the top 100 influencers aggregated across all news media categories in 2020, seventy-five percent were not present in 2016, demonstrating how difficult it is to retain influencer status. The number of influencers affiliated with media organizations declined by 10% between 2016 and 2020, replaced by influencers affiliated with political organizations with center or right orientations and those with independent organizational affiliation. Most of the influencers who appeared in 2020 were associated with prominent media or political party organizations.

Future research should build on this structural analysis by examining the content of the messages. Content analysis is needed to distinguish between tweets that are positively and negatively quoted and to develop measures of influence that go beyond the ability to attract attention from retweeters. For example, an urgent question to answer is whether the influence of unaffiliated Twitter influencers goes beyond being news spreaders: do they also have the ability to set the issue agenda? Our study is limited to describing what happened on Twitter. Future research should analyze message content for clues about the ability of influencers to mobilize voters and social movements offline. We also focused on the flow of information from influencers to those who retweet them. Future research should investigate how the actions of the retweeters and followers affect the influencers, how influencers form networks across types of media, and what are the offline consequences of polarization of Twitter influencers and users, including the impact on voting. It should also be possible to monitor interactions on other social media and during non-election periods to permit finer grained analysis of the new entrants.

4.4 Methods

4.4.1 News Media URL Classification

The website `www.allsides.com` (AS) rates media bias using a combination of several methods such as blind surveys, editorial review, third party analysis (e.g. academic research), independent review and community feedbacks (see `www.allsides.com/media-bias/media-bias-\rating-methods` for a detailed explanation of their methodology). The website `mediabiasfact\check.com` (MBFC) scores media bias by evaluating wording, sourcing, and story choices as well as political endorsement (see `mediabiasfactcheck.com/methodology`). MBFC is maintained by a small independent team of researchers and journalists, offers the largest set of biased and inaccurate news sources among five fact checking datasets [116], and is widely used for labeling bias and veracity of news sources (e.g., in [117]–[119]).

To be consistent with the results from 2016 [18], we discard as insignificant outlets that accumulate less than 1% of the cumulative number of tweets of the more popular outlets in each category. Removing uniformly insignificant outlets from all categories also ensures that the tweet volume in each category is independent of the number of outlets classified in this category by AS and MBFC. The full lists of outlets in each category in 2016 and 2020 are given in Tabs. A.1 and A.2. AS and MBFC updated their bias classification for several outlets between 2016 and 2020, changing the classification used in our analyses as well. For example, CNN Web News was classified in the *center* category in 2016 by AS and then in the *left-leaning* category in 2020, reflecting a bias shift occurring during this time (see `www.allsides.com/blog/yes-cnns-media-bias-has-shifted-left`).

In Ref. [18], the fake news and extreme-bias categories were based on the classification of a team of media experts (available at `github.com/alexbovet/opensources`) and was cross-checked using the factual reporting scores from MBFC. As the classification source from 2016 was not updated in 2020, we use the list of outlets classified as "questionable sources" from MBFC as a reference for 2020. MBFC describes a questionable source as one "that exhibits one or more of the following: extreme bias, consistent promotion of propaganda/conspiracies, poor or no sourcing to credible information, a complete lack of transparency and/or is fake news." MBFC rates the factual reporting of each source on a scale from 0 (very high) to 10 (very low) based on their history of reporting factually and backing up claims with well-sourced evidence. Outlets with a level of "low" (score of 7, 8 or 9) or "very low" (score of 10) are classified in the fake news category while outlets with a "mixed" level (score of 5 or 6) are classified in the extremely biased category. No outlets in the disinformation categories have a level higher than "mixed." A "low" or "very low" factual reporting level on MBCF corresponds to sources that rarely, or almost never use credible sources and "need to be checked for intentional fake news, conspiracy, and propaganda." A "mixed" level is assigned to sources that "do not always use proper sourcing or source to other biased/mixed factual sources." We also verify that all outlets in the extremely biased category have a "bias" reported on MBFC of "right", "extreme right", "left" or "extreme left."

We identify in our datasets (we give the top hostname as an example in parenthesis) for the fake news category: 16 hostnames in 2016 (top: `thegatewaypundit.com`) and 20 hostnames in 2020 (top: `thegatewaypundit.com`), for the extremely biased (right) category: 17 hostnames in 2016 (top: `breitbart.com`) and 10 hostnames in 2020 (top: `breitbart.com`), for the extremely biased (left) category: 7 hostnames in 2016 (top: `dailynewsbin.com`) and 7 hostnames in 2020 (top: `occupydemocrats.com`), for the left news category: 18 hostnames in 2016 (top: `huffingtonpost.com`) and 18 hostnames in 2020 (top: `rawstory.com`), for the left-leaning news category: 19 hostnames in 2016 (top: `nytimes.com`) and 19 hostnames in 2020 (top: `nytimes.com`), for the center news category: 13 hostnames in 2016 (top: `cnn.com`) and 13 hostnames in

2020 (top: `thehill.com`), for the right-leaning news category: 7 hostnames in 2016 (top: `wsj.com`) and 13 hostnames in 2020 (top: `nypost.com`), for right news category: 20 hostnames in 2016 (top: `foxnews.com`) and 19 hostnames in 2020 (top: `foxnews.com`). The full lists of outlets in each category in 2016 and 2020 are given in SI Tabs. A.1 and A.2.

4.4.2 Influencer Type Classification

For each of the years 2016 and 2020, we manually classified the top-25 influencers in each news media category as affiliated to a media organization or a political organization, or unaffiliated (classified either as an independent user or as an unidentified "other" user). The manual labeling procedure was as follows: Eight of the authors were randomly assigned a subset of the union of the top-25 influencers in these category lists to independently classify, such that each subset was examined by three different authors. Each author was shown the account name of the influencer along with descriptions, posts, and all available non-Twitter information such as their Wikipedia entry. Each influencer was then assigned their category based on the majority vote of the three independent classifications.

4.4.3 Similarity Network Analysis

We start by creating for each influencer i a vector $\vec{S^i}$ of size U, which stands for the number of users in our dataset. We used a set of 588 influencers for the 2016 dataset, and a set of 661 influencers for the 2020 dataset. An index u is assigned to the specific user. The vector element s_u^i defines the number of times user u has retweeted influencer i. Then, we create the adjacency matrix $\mathbf{A}$ of size $I \times I$ for our similarity networks by setting $a_{i1,i2}$ to the cosine similarity between vectors $\vec{S^{i_1}}$ and $\vec{S^{i_2}}$. It follows that the higher the cosine similarity, the more users have the similar number of retweets for influencers i_1, i_2.

We detect communities in the similarity network using the Louvain algorithm [45]. In the similarity networks for both election years, we found two communities. Using the accounts of influencers in each community, we found that both election years one community contains influencers primarily associated with fake and right-biased news categories, while the other contains influencers from center and left-biased news categories. This split coincides with an underlying division among the Twitter user bases in the content they propagate.

We quantify the severity of this split using two measures of separation between communities. First is modularity that computes the sum of difference between the fraction of edges within each community and such fraction expected within this community in a random network with the same number of nodes and edges. This metric has a range of $[-0.5, 1]$ [46]. A positive value indicates the presence of communities separated from each other. The closer the modularity is to 1, the stronger communities are separated.

The second measure uses the normalized cut, which is the sum of the weights of every edge that links a pair communities divided by the sum of the weights of all edges. The result has a range of $[0, 1]$ where the smaller the value, the stronger the separation among communities.

4.4.4 Latent Ideology Estimation

The latent ideology estimation follows the method developed in [12], [47] adapted to using retweet interactions instead of following relations. As in [12], we use correspondence analysis [48] (CA) to infer ideological positions of Twitter users.

The adjacency matrix, $\mathbf{A}$, of the retweet network between the influencers and their retweeters is the matrix with element a_{ij} equal to the number of times user i retweeted influencer j. We only select tweets that have been sent from the official Twitter client in order to limit the presence of bots and professional accounts and we also remove users that show a low interest in the U.S. elections by removing users that retweeted less than three different influencers. For the 2016 data, the matrix $\mathbf{A}$ has 751,311 rows corresponding to distinct users, 593 columns corresponding to influencers and the total number of retweets equal to 39,385,772. For the 2020 data, the matrix $\mathbf{A}$ has 2,034,970 rows corresponding to distinct users, 591 columns corresponding to influencers and the total number of retweets equal to 153,463,788.

The CA method is executed in the following steps [49]. The matrix of standardized residuals of the adjacency matrix is computed as $\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{rc})\mathbf{D}_c^{-1/2}$, where $\mathbf{P} = \mathbf{A}(\sum_{ij} a_{ij})^{-1}$ is the adjacency matrix normalized by the total number of retweets, $\mathbf{r} = \mathbf{P1}$ is the vector of row sums, $\mathbf{c} = \mathbf{1}^T\mathbf{P}$ is the vector of column sums, $\mathbf{D}_r = \text{diag}(\mathbf{r})$ and $\mathbf{D}_c = \text{diag}(\mathbf{c})$. Using the standardized residuals allows the inference to account for the variation of popularity and activity of the influencers and the users, respectively [12]. Then, a SVD is computed such that $\mathbf{S} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^T$ with $\mathbf{U}\mathbf{U}^T = \mathbf{V}\mathbf{V}^T = \mathbf{I}$ and $\mathbf{D}_\alpha$ being a diagonal matrix with the singular values on its diagonal. The positions of the users are given by the standard row coordinates: $\mathbf{X} = \mathbf{D}_r^{-1/2}\mathbf{U}$ where we only consider the first dimension, corresponding to the largest singular value. Finally, the ideological positions of the users are found by standardizing the row coordinates to have a mean of zero and a standard deviation of one. The ideological position of the influencers is given by the median of the weighted positions of their retweeters.

We tested the robustness of our method by varying the way we construct matrix $\mathbf{A}$ as follow: 1) removing entries with weight 1 to discard relations showing a weak ideological alignment; 2) considering the logarithm of the number of retweets as weight for influencer for a sublinear relation between the number of retweets and the strength of ideology alignment; 3) considering a random subsample of the 2020 retweet data of the same size than the 2016 retweet data to control for a potential effect of the difference in sizes of the two datasets. All of these robustness tests match the results of our initial method with correlation coefficients between the user position distributions in the robustness tests and in the initial configuration at above 0.995. We also compare the users' latent ideology distribution with the users average leaning distribution and find a correlation above 0.90 for 2016 and 2020. The average leaning of users is computed for all users having at least three tweets classified in at least one news media category and estimated as the weighted average of the news media category positions, given as: fake news = 4/3, extreme-right bias = 1, right = 2/3, right-leaning = 1/3, center = 0, left-leaning = -1/3, left = -2/3, extreme-left bias = -1.

CHAPTER 5

EVALUATION OF THE ECHO CHAMBER IMPACT ON SOCIAL MEDIA

Previous chapter analyzes the changes between 2020 and 2016 U.S. presidential elections. One of the main results of the study is the increase in polarization from 2016 to 2020: respect to 2016, in 2020 the two factions interacted less and increased the distance between their political position. Our finding is also supported by recent events such as the Capitol Hill riot, which could be seen as an expression of such increase in polarization and radicalization of the electorate.

Although users segregation is the consequence of a plethora of contributing causes, online environmental factors may foster polarization and favor the raise of online echo chambers. In particular, platforms' feed algorithms may be responsible for increasing the homogeneity of users' news diet and boost homophilic interactions. In this chapter, we provide a comparative analysis of the echo chamber effect in four different social media around several topics. By means of networks built upon users' interaction, we also compare the spreading dynamic and study the presence of echo chambers around the same topic in different platforms. Our analysis provide insights on the differences and similarities of echo chambers across different social media and around several topics.

5.1 Introduction

Social media radically changed the mechanism we access information and form our opinions [2], [11], [120]–[122]. We need to understand how people seek or avoid information and how those decisions affect their behavior [123], especially when the news cycle—dominated by the disintermediated diffusion of information—alters the way information is consumed and reported on. A recent study [14] limited to Twitter claimed that fake news travels faster than real ones. However, a multitude of factors affects information spreading on social media platforms. Online polarization, for instance, may foster misinformation spreading [2], [68]. Our attention span remains limited [4], [124], and feed algorithms might limit our selection process by suggesting contents similar to the ones we are usually exposed to [3], [5], [119]. Furthermore, users show the tendency to favor information adhering to their beliefs and join groups formed around a shared narrative, i.e., echo chambers [2], [6]–[10]. We can broadly define echo chambers as environments in which the opinion, political leaning, or belief of users about a topic get reinforced due to repeated interactions with peers or sources having similar tendencies and attitudes. Selective exposure [125] and confirmation bias [126] (i.e., the tendency to seek information adhering to pre-existing opinions) may explain the emergence of echo chambers on social media [2], [9], [40], [57].

According to group polarization theory [42], an echo chamber can act as a mechanism to reinforce an existing opinion within a group, and as a result, move the entire group towards more extreme positions. Echo chambers have been shown to exist in various forms of online media such as blogs [127], forums [128], and social-media sites [12], [129], [130]. Some studies point out echo chambers as an emerging effect of human tendencies, such as selective exposure, contagion, and group polarization [5], [42], [131]–[133]. However, recently, the effects and the very existence of echo chambers have been questioned [11]–[13]. This issue is also fueled by the scarcity of comparative studies on social media, especially for what concerns news consumption [43]. In this context, the debate around echo chambers is fundamental to understanding social media's influence on information consumption and public opinion formation. In this paper, we explore the key differences between social media platforms and how they are likely to influence the formation of echo chambers or not. As recently shown in the case of selective exposure to news outlets, studies considering multiple-platforms can offer a fresh view to long-debated problems [134]. Different platforms offer different interaction paradigms to users, ranging from retweets and mentions on Twitter to likes and comments in groups on Facebook, thus triggering very different social dynamics [135]. We introduce an operational definition of echo chambers to provide a common methodological ground to explore how different platforms influence their formation. In particular, we operationalize the two common elements that characterize echo chambers into observables that can be quantified and empirically measured, namely: (i) the inference of the user's leaning for a specific topic (e.g., politics, vaccines), (ii) the structure of their social interactions on the platform. Then, we use these elements to assess echo chambers' presence by looking at two different aspects: (i) homophily in interactions concerning a specific topic and (ii) bias in the information diffusion from like-minded sources. We focus our analysis on multiple platforms: Facebook, Twitter, Reddit, and Gab. These platforms present similar features and functionalities (e.g., they all allow social feedback actions such as likes or upvotes) and design (e.g., Gab is similar to Twitter) but also distinctive features (e.g., Reddit is structured in communities of interest called subreddits). Reddit is one of the most visited websites worldwide[1] and is organized as a forum to collect discussions on a wide range of topics, from politics to emotional support. Gab claims to be a social platform aimed at protecting freedom of speech. However, low moderation and regulation on content has resulted in widespread hate speech. For these reasons, it has been repeatedly suspended by its service provider, and its mobile app has been banned from both App and Play stores [136]. Overall, we account for the interactions of more than 1M active users

[1] https://www.alexa.com/siteinfo/reddit.com

on the four platforms, for a total of more than 100M unique pieces of content, including posts and social interactions. Our analysis shows that platforms organized around social networks and news feed algorithms, such as Facebook and Twitter, favor the emergence of echo chambers.

We conclude this work by directly comparing news consumption on Facebook and Reddit, finding higher segregation on Facebook than on Reddit.

5.2 Characterizing Echo Chambers in Social Media

5.2.1 Operational Definitions

To explore the key differences between social media platforms and how they influence echo chambers' formation, we need to operationalize a definition for them. First, we need to identify the attitude of users at a micro-level. On online social media, the *individual leaning* of a user i toward a specific topic, x_i, can be inferred in different ways, via the content produced, or the endorsement network among users [137]. Concerning content, we can define the leaning as the attitude expressed by a piece of content towards a specific topic. This leaning can be explicit (e.g., arguments supporting a narrative) or implicit (e.g., framing and agenda-setting). Let us consider a user i producing a number a_i of contents, $\mathcal{C}_i = \{c_1, c_2, \dots, c_{a_i}\}$, where a_i is the *activity* of user i and each content leaning is assigned a numeric value. Then the individual leaning of user i can be defined as the average of the leanings of produced contents,

$$x_i \equiv \frac{\sum_{j=1}^{a_i} c_j}{a_i}. \tag{5.1}$$

Once inferred individual leanings, *polarization* can be defined as a state of the system such that the distribution of leanings, $P(x)$, is concentrated in one or more clusters. A possible example is the case of a single cluster, distinguishable by a single, extreme peak in $P(x)$. Another example is the typical case of topics characterized by positive versus negative stances, in which a bimodal distribution can describe polarization. For instance, if opinions are assumed to be embedded in a one-dimensional space [138], $x \in [-1, +1]$ without loss of generality, as usual for controversial topics, then polarization is characterized by two well-separated peaks in $P(x)$, for positive and negative opinions. In contrast, neutral ones are absent or underrepresented in the population. Note that polarization can happen independently from the structure or the very presence of social interactions. *Homophily* in social interactions can be quantified by representing interactions as a social network and then analyzing its structure concerning the opinions of the users [10], [105], [139]. Social networks can be reconstructed in different ways from online social media, where links represent social relationships or interactions. Since we are interested in capturing the possible exchange of opinions between users, we assume links as the substrate over which information may flow. For instance, if user i follows user j on Twitter, user i can see tweets produced by user j, there is a flow of information from node j to node i in the network. When the reconstructed network is directed, we assume the link direction points to potential influencers (opposite of information flow). Actions such as mentions or retweets may convey similar flows. In some cases, direct relations between users are not available in the data, so one needs to assume some proxy for social connections, e.g., a link between two users if they comment on the same post on Facebook. Crucially, the two elements characterizing the presence of echo chambers, polarization, and homophilic interactions, should be quantified independently.

5.2.2 Implementation on Social Media

This section explains how we implement the operational definitions defined above on different social media. For each medium, we detail (i) how we quantify users' leaning, and (ii) how we reconstruct how the information spread.

Twitter. We consider the set of tweets posted by user i that contain links to news outlets of known political leaning. Each news outlet is associated with a political leaning score ranging from extreme left to extreme right following the Materials and Methods classification. We infer the individual leaning of a user i, $x_i \in [-1, +1]$ by averaging the news organizations' scores linked by user i according to (5.1). We analyze three different data sets collected on Twitter related to controversial topics: gun control, Obamacare, and abortion. For each data set, the social interaction network is reconstructed using the following relation so that there is a direct link from node i to node j if user i follows user j (i.e., the source). Henceforth we focus on the data set about abortion, and others are reported in the results section without discussion.

Facebook. We quantify the individual leaning of users considering endorsements in the form of likes to posts. Posts are produced by pages that are labeled in a certain number of categories, and to each category, we assign a numerical value (e.g., Anti-Vax (+1) or Pro-Vax (-1)). Each like to a post (only one like per post is allowed) represents an endorsement for that content, which is assumed to be aligned with the leaning associated with the page. Thus, the user's leaning is defined as the average of the content leanings of the posts liked by the user, according to (5.1).

We analyze three different data sets collected on Facebook regarding a specific topic of discussion: vaccines, science versus conspiracy, and news. The interaction network is defined by considering comments. In such an interaction network, two users are connected if they co-commented at least one post. Henceforth we focus on the data set about vaccines and news, and others are reported in the results section without discussion.

Reddit. The individual leaning of users is quantified similarly to Twitter by considering the links to news organizations in the content produced by the users, submissions, and comments. We build the interaction network considering comments and submissions. There exists a direct link from node i to node j if user i comments on a submission or comment by user j (we assume that i reads the comment they are replying to, which is written by j).

We analyze three data sets collected on different subreddits: the_donald, politics, and news. In the following, we focus on the data set collected on the Politics and the News subreddit, and others are reported in the results section without discussion.

Gab. The political leaning x_i of user i is computed by considering the set of contents posted by user i containing a link to news outlets of a known political leaning, similarly to Twitter and Reddit. To obtain the leaning x_i of user i, we averaged the scores of each link posted by user i according to (5.1). The interaction network is reconstructed by exploiting the co-commenting relationships under posts in the same way as for Facebook. Given two users i and j, an undirected edge between i and j exists if and only if they comment under the same post.

5.3 Comparative Analysis

In the following, we perform a comparative analysis of four different social media. We select one dataset for each social media: Abortion (Twitter), Vaccines

(Facebook), Politics (Reddit), and Gab as a whole. Results for other datasets for the same medium are qualitatively similar and we show them in section 5.5.5. We first characterize echo chambers in the networks' topology, then look at their effects on information diffusion. Finally, we directly compare news consumption on Facebook and Reddit.

5.3.1 Polarization and Homophily in the Interaction Networks

The network's topology can reveal echo chambers, where users are surrounded by peers with similar leaning, and thus they get exposed with a higher probability to similar contents. In network terms, this translates into a node i with a given leaning x_i more likely to be connected with nodes with a leaning close to x_i [10]. This concept can be quantified by defining, for each user i, the average leaning of their neighborhood, as $x_i^N \equiv \frac{1}{k_i^{\rightarrow}} \sum_j A_{ij} x_j$, where A_{ij} is the adjacency matrix of the interaction network, $A_{ij} = 1$ if there is a link from node i to node j, $A_{ij} = 0$ otherwise, and $k_i^{\rightarrow} = \sum_j A_{ij}$ is the out-degree of node i. Fig. 5.1 shows the correlation between the leaning of a user i and the leaning of their neighbors, x_i^N, for the four social media under consideration. The probability distributions $P(x)$ (individual leaning) and $P^N(x)$ (average leaning of neighbors) are plotted on the x and y axis, respectively. All plots are color-coded contour maps, representing the number of users in the phase space (x, x^N): the brighter the area in the plan,

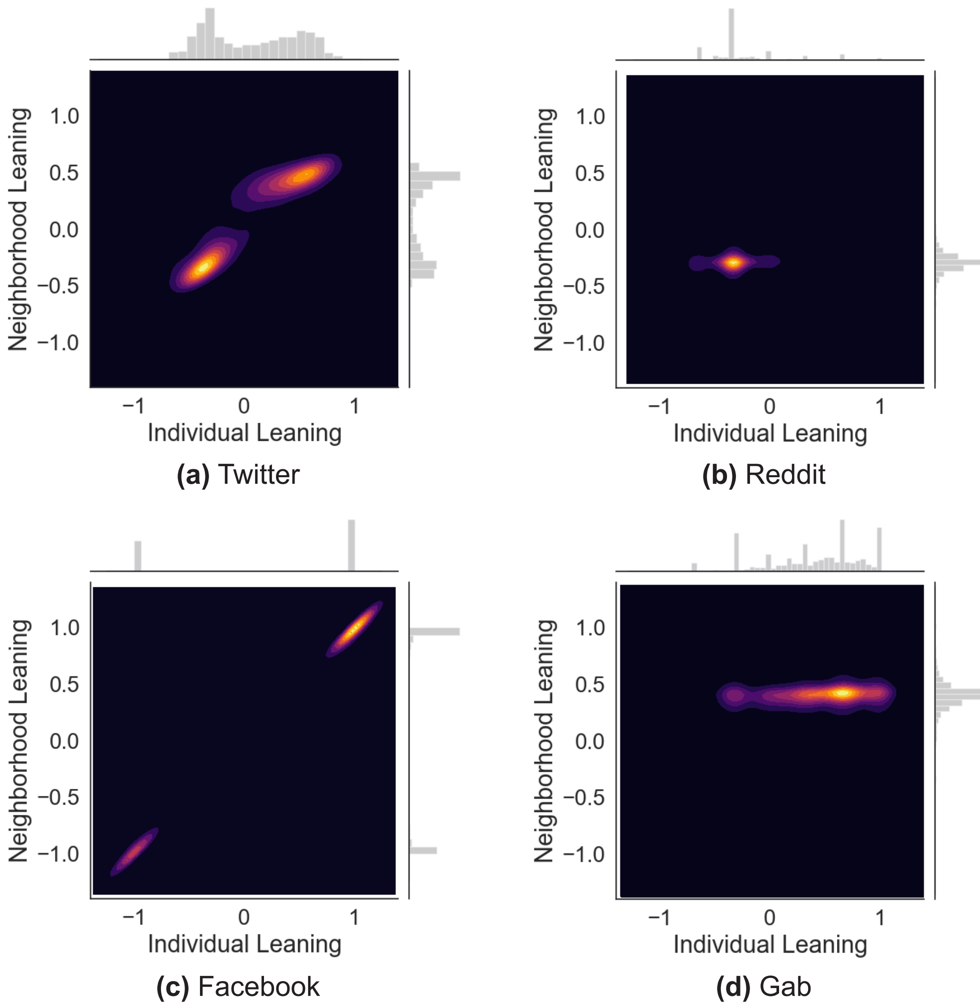


Fig. 5.1: Joint distribution of the leaning of users x and the average leaning of their neighborhood x^{NN} for different data sets. Colors represent the density of users: the lighter, the larger the number of users. Marginal distribution $P(x)$ and $P^N(x)$ are plotted on the x and y axis, respectively. Facebook and Twitter present by homophilic clustering.

the larger the density of users in that area. The topics of vaccines and abortion, on Facebook and Twitter, respectively show a strong correlation between the leaning of a user and the average leaning of their nearest neighbors. Similar behavior is found for different topics from the same social media platform, as shown in section 5.5.5. Conversely, Reddit, and Gab show a different picture. The corresponding plots in Fig. 5.1 display a single bright area, indicating that users do not split into groups with opposite leaning but form a single community, biased to the left (Reddit) or the right (Gab). Similar results are found for different data sets on Reddit, as shown in section 5.5.5

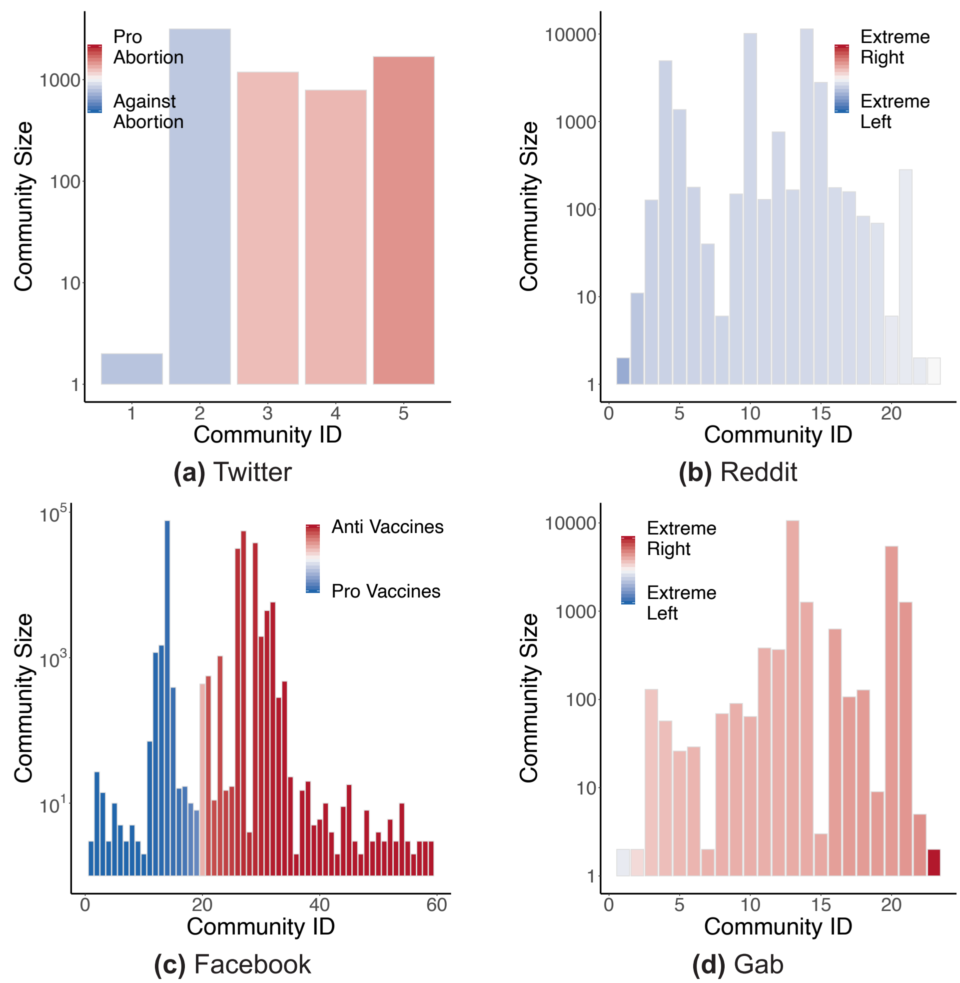


Fig. 5.2: Size and average leaning of communities detected in different data sets. Panels a and c show the full spectrum of leanings related to the topics of abortions and vaccines w.r.t communities in panels b and d where the political leaning is less sparse.

The presence of homophilic interactions can be confirmed by the community structure of the interaction networks. We detect communities by applying the Louvain algorithm [45], removing singleton communities with only one user. Then, we computed each community's average leaning, determined as the average of individual leanings of its members. Fig. 5.2 shows the communities emerging for each social medium, arranged by increasing average leaning on the x-axis (color-coded from blue to red), while the y-axis reports the size of the community. On Facebook and Twitter, communities span the whole spectrum of possible leanings, but users with similar leanings form each community. Some communities are characterized by a robust average leaning, especially in the case of Facebook. These results are in accordance with the observation of homophilic interactions. Instead, communities on Reddit and Gab do not cover the whole spectrum, and all show similar average leaning. Furthermore, the almost total absence of communities with leaning very close to 0 confirms the polarized state of the systems.

5.3.2 Effects on Information Spreading

Simple models of information spreading can gauge the presence of echo chambers: users are expected to exchange information more likely with peers sharing a similar leaning [10], [44], [140]. Classical epidemic models such as the susceptible-infected-recovered (SIR) model [141] have been used to study the diffusion of information, such as rumors or news [142]–[144]. In the SIR model, each agent can be in either of three states: susceptible (unaware of the circulating information), infectious (aware and willing to spread it further), or recovered (knowledgeable but not ready to transmit it anymore). Susceptible (unaware) users may become infectious (aware) upon contact with infected neighbors, with a specific transmission probability β. Infectious users can spontaneously become recovered with probability ν. To measure the effects of the leaning of users on the diffusion of information, we run the SIR dynamics on the interaction networks, by starting the epidemic process with only one node i infected, and stopping it when no more infectious nodes are left.

The set of nodes in a recovered state at the end of the dynamics started with user i as a seed of infection, i.e., those that become aware of the information initially propagated by user i forms the *set of influence* of user i, $\mathcal{I}_i$ [50]. Thus, the set of influence of a user represents those individuals that can be reached by a piece of content sent by him/her, depending on the effective infection ratio β/ν. One can compute the average leaning of the set of influence of user i, μ_i, as

$$\mu_i \equiv |\mathcal{I}_i|^{-1} \sum_{j \in \mathcal{I}_i} x_j. \tag{5.2}$$

The quantity μ_i indicates how polarized are the users that can be reached by a message initially propagated by user i [10].

Fig. 5.3 shows the average leaning $\langle \mu(x) \rangle$ of the influence sets reached by users with leaning x, for the different data sets under consideration. The recovery rate ν is fixed at 0.2 for every dataset. In contrast, the ratio between the infection rate β and average degree $\langle k \rangle$ depends on the specific dataset and is reported in the caption of each figure.

Again, one can observe a clear distinction between Facebook and Twitter, on one side, and Reddit and Gab on the other side. For the topics of vaccines and abortion, on Facebook and Twitter, respectively, users with a given leaning are much more likely to be reached by information propagated by users with similar leaning, i.e., $\langle \mu(x) \rangle \sim x$. Similar behavior is found for different topics from the same social media platform, as shown in section 5.5.5. Conversely, Reddit and Gab show a different behavior: the average leaning of the set of influence, $\langle \mu(x) \rangle$, does not depend on the leaning x. As expected, the average leaning in these media is not zero. Still, it assumes negative (positive) values in Reddit (Gab), indicating that the users of this platform are more likely to receive left (right)-leaning content.

These results indicate that information diffusion is biased toward individuals who share similar leaning in some social media, namely Twitter and Facebook. In contrast, in others – Reddit and Gab in our analysis – this effect is absent. Such a latter configuration may depend upon two factors: a) Gab and Reddit are not bursting the echo chamber effects, or b) we are observing the dynamic inside a single echo chamber.

Our results are robust for different values of the effective infection ratio β/ν, as shown in section 5.5.6 Furthermore, Fig. 5.3 shows that the spreading capacity, represented by the average size of the influence sets (color-coded in Fig. 5.3), depends on the leaning of the users. On Twitter, pro-abortion users are more likely to reach larger audiences. The same is true for anti-vax users on Facebook, left-leaning users on Reddit, and right-leaning users on Gab (in this data set, left-leaning users are almost absent).

5.3.3 News Consumption on Facebook and Reddit

The striking differences observed across social media, in terms of homophily in the interaction networks and information diffusion, could be attributed to different topics taken into account. For this reason, here we compare Facebook and Reddit on a common topic, news consumption. Facebook and Reddit are particularly apt to a cross-comparison since they share the definition of individual leaning (computed by using the classification provided by mediabiasfactcheck.org, see Materials and Methods for further details) and the rationale in creating connections among users that is based on an interaction network. Fig. 5.4 shows a direct comparison of news consumption on Facebook and Reddit along the metrics used in the previous Sections to quantify the presence of echo chambers: i) the corre-

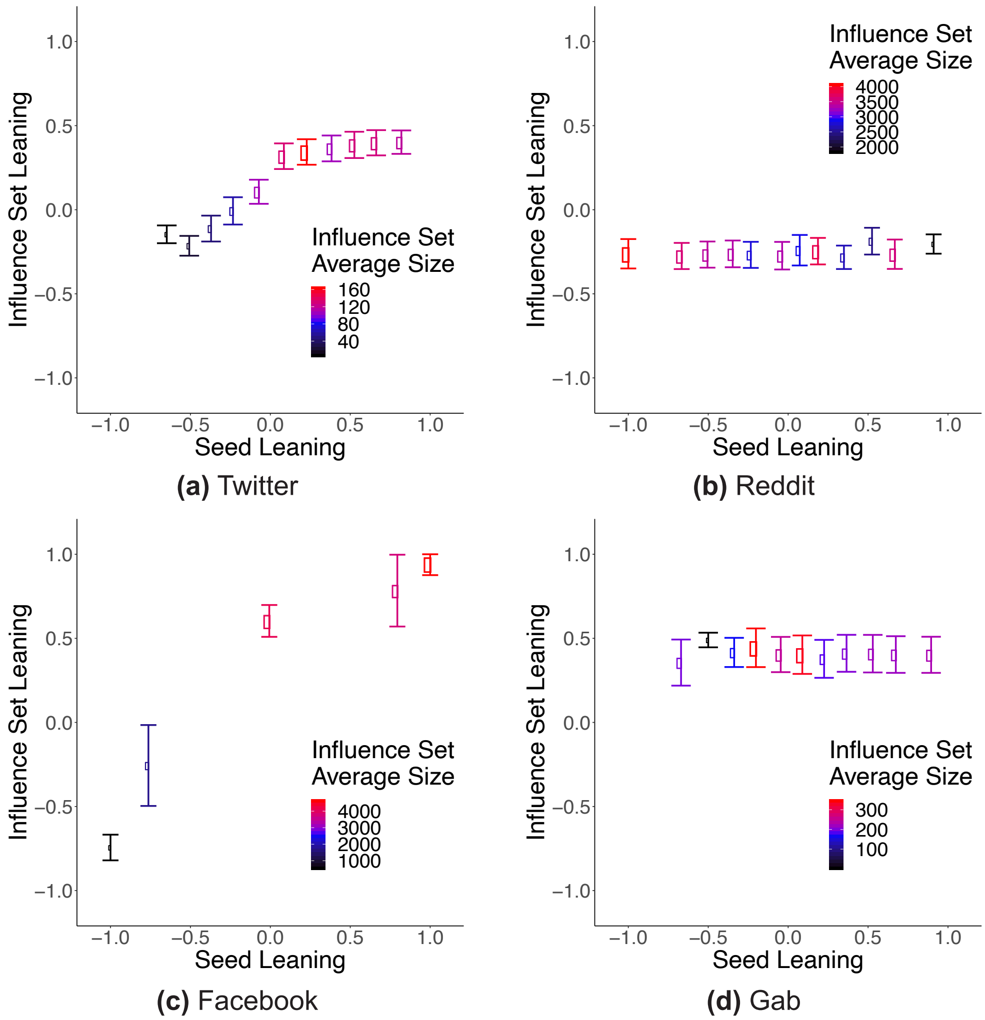


Fig. 5.3: Average leaning $\langle\mu(x)\rangle$ of the influence sets reached by users with leaning x, for different data sets under consideration. Size and color of each point represent the average size of the influence sets. The parameters of the SIR dynamics are set to $\beta = 0.10\langle k\rangle^{-1}$ for panel (a), $\beta = 0.01\langle k\rangle^{-1}$ for panel (b), $\beta = 0.05\langle k\rangle^{-1}$ for panel (c) and $\beta = 0.05\langle k\rangle^{-1}$ for panel (d), while ν is fixed at 0.2 for all simulations.

lation between the leaning of a user x and the average leaning of neighbors x^N (top row), ii) the average leaning of communities detected in the networks (middle row), and iii) the average leaning $\langle\mu(x)\rangle$ of the influence sets reached by users with leaning x, by running SIR dynamics (bottom row). One can see that all three measures confirm the picture obtained for other data sets: On Facebook, we observe a clear separation among users depending on their leaning, while on Reddit, users' leanings are more homogeneous and show only one peak. In the latter social media, even users displaying a more extreme leaning (noticeable in the marginal histogram of Figure 5.4 column (b) top row) tend to interact with the majority. Moreover, on Facebook, the seed user's leaning affects who the final recipients of the information are, therefore indicating the presence of echo chambers. On Reddit, this effect is absent.

5.4 Conclusions

Social media platforms provide direct access to an unprecedented amount of content. Platforms originally designed for user entertainment changed the way information spread. Indeed, feed algorithms mediate and influence the content promotion accounting for users' preferences and attitudes. Such a paradigm shift affected the construction of social perceptions and the framing of narratives; it may influence policy-making, political communication, and the evolution of public debate, especially on polarizing topics. Indeed, users online tend to prefer information adhering to their worldviews, ignore dissenting information, and form polarized groups around shared narratives. Furthermore, when polarization is high, misinformation quickly proliferates.

Some argued that the veracity of the information might be used as a determinant for information spreading patterns. However, selective exposure dominates content consumption on social media, and different platforms may trigger very different dynamics. In this work, we explore the key differences between the leading social media platforms and how they are likely to influence the formation of echo chambers and information spreading.

To assess the different dynamics, we perform a comparative analysis on more than 100M pieces of content concerning controversial topics (e.g., gun control, vaccination, abortion) from Gab, Facebook, Reddit, and Twitter. The analysis focuses on two main dimensions: i) homophily in the interaction networks and ii) bias in the information diffusion toward like-minded peers. Our results show that the aggregation in homophilic clusters of users dominate online dynamics. However, a direct comparison of news consumption on Facebook and Reddit shows higher segregation on Facebook. Furthermore, we find significant differences across platforms in terms of homophilic patterns in the network structure and biases in the information diffusion towards like-minded users. A clear-cut distinction emerges between social media having a feed algorithm tweakable by the users (e.g., Reddit) and social media that don't provide such an option (e.g., Facebook and Twitter). Our work provides important insights into the understanding of social dynamics and information consumption on social media. The next envisioned step addresses the temporal dimension of echo chambers to understand better how different social feedback mechanisms, specific to distinct platforms, can impact their formation.

5.5 Materials and Methods

Here we provide details about the labelling of news outlets and the data sets considered.

5.5.1 Labelling of Media Sources

The labelling of news outlets is based on the information provided by Media Bias/Fact Check (MBFC `https://mediabiasfactcheck.com`), an independent fact-checking organization that rates news outlets on the base of the reliability and of the political bias of the contents they produce and share. The website provides the political bias related to a wide range of media outlets. The labelling provided by MBFC, retrieved in June 2019, ranges from Extreme Left to Extreme Right for what concerns the political bias. Moreover, certain media outlets are classified as 'questionable' sources or 'conspiracy-pseudoscience' sources if they tend to publish misinformation or false contents. Often, such news outlets (without an explicit political label reported by MBFC) actually display a political bias that is reported in their description, as shown in Figure 5.5.

Considering the importance of including such media outlets in our analysis, we manually reported their classification from the description provided by MBFC, thus adding 468 outlets to the pool of 1722 news outlets that already have a clear political label. The total number of labelled news outlets is 2190 and the overall leaning is summarized in Figure 5.6. In order to compute the individual leaning of users we convert each label into a numerical value, namely, -1 for Extreme Left, -0.66 for Left, -0.33 for Left-Center, 0 for Least Biased, 0.33 for Right-Center, 0.66 for Right and +1 for Extreme Right.

5.5.2 Data Availability Statement

For what concerns Gab, all data are available on the Pushshift public repository (`https://pushshift.io/what-is-pushshift-io/`) at this link `https://files.pushshift.io/gab/`. Reddit Data are available on the Pushshift public repository at this link `https://search.pushshift.io/reddit/`. For what concerns Facebook and Twitter, we provide data according to their Terms of Services on the corresponding author institutional page at this link `https://walterquattrociocchi.site.uniroma1.it/ricerca`. For news outlet classification, we used data from Media Bias Fact-check (`https://mediabiasfactcheck.com`), an independent fact-checking organization. For further details about data, refer to the following section.

5.5.3 Empirical Data Sets

Table 5.1 reports summary statistics of the data sets under consideration. Due to the structural differences among platforms, each dataset has different features. For Twitter, we used tweets regarding three topics collected by Garimella et al. [8], namely **Gun control**, **Obamacare**, and **Abortion**. Tweets linking to a news source with a known bias are classified based on MBFC. Facebook data sets were created by using Facebook Graph API and were previously explored in [107] (**Science and Conspiracy**), [145] (**Vaccines**) and [3] (**News**). For the two data sets **Science and Conspiracy** and **Vaccines**, data were labelled in a binary way, respectively pro vaccines/anti vaccines and pro science/conspiracy based on the page they were posted. Posts in the data set **News** were instead classified based on MBFC labelling. Reddit datasets have been obtained by downloading comments and submission posted in the subreddit **Politics**, **The Donald** and **News** and labelled according to the classification obtained from MBFC. **Gab** data set has been collected from `https://files.pushshift.io/gab` and contains posts, replies and quotes. Posts were labelled according to MBFC classification. We provide a detailed description of each datasets in the next section.

5.5.4 Dataset Detailed Description

Twitter

We follow a two-step procedure for creating the Twitter datasets. First, tweets during the interest periods are retrieved from the Internet Archive Twitter Stream.[2] For each topic, we use the keywords specified by Lu et al. [146]. Each user that has posted 5 or more tweets on the topic during the window of interest is considered active. We then use the Twitter's REST API[3] to collect all tweets and followers for each active user. These tweets and relationships are the basis for reconstructing each network. For more info on the datasets, see the work by Garimella et al. [8].

Gun control: The interest window spans 14 days in June 2016. We consider $C = 19$M tweets produced by $N = 7506$ users. We reconstruct a directed follow network formed by $E = 1\,053\,275$ directed edges. The largest weakly connected component includes more than 99% of nodes. We identify the individual leaning of $N_c = 6994$ users.

Obamacare: The interest window spans 7 days in June 2016. We consider $C = 34$M tweets produced by $N = 8773$ users. We reconstruct a directed follow network formed by $E = 3\,797\,871$ directed edges. The largest weakly connected component includes more than 99% of nodes. We identify the individual leaning

[2] `https://archive.org/details/twitterstream`
[3] `https://developer.twitter.com/en/docs/twitter-api/v1`

of $N_c = 7899$ users.

Abortion: The interest window spans 7 days in June 2016. We consider $C = 34$M tweets produced by $N = 3995$ users. We reconstruct a directed follow network formed by $E = 2\,330\,276$ directed edges. The largest weakly connected component includes more than 99% of nodes. We identify the individual leaning of $N_c = 3809$ users.

Facebook

Science and Conspiracy: The dataset was built by downloading posts of selected Facebook pages divided into two groups, namely conspiracy news and science news. Conspiracy pages were selected based on their name, their self description and with the aid of debunking pages. The selection process was iterated until convergence among annotators. The dataset, that includes post from pages and comments to such posts, was created by using Facebook Graph API and has been previously explored [107]. We consider $75\,172$ posts by 73 pages categorized in Science (34) and Conspiracy (39) that involve $N = 183\,378$ active users (at least 1 like and 1 comments), for which we identify the individual leaning, that co-commented $20\,807\,976$ times. Using this dataset we build an undirected network, where two users (nodes) are connected if and only if they commented under the same post at least once. The largest connected component of the co-commenting network has $G = 181\,960$ nodes and $E = 20\,807\,491$ links.

Vaccines: The dataset was generated in three steps: first a search for pages containing the keywords vaccine, vaccines, or vaccination was made. Then the raw outcome was cleaned from spurious pages. Finally, all the posts and comments of selected pages were downloaded and pages were manually classified in Pro-Vax and Anti-Vax groups. The dataset was created by using Facebook Graph API and has been previously explored [145]. We consider $94\,776$ posts by 243 pages categorized in Pro-Vax (145) and Anti-Vax (98) that involve $221\,758$ active users (at least 1 like and 1 comment), for which we identify the individual leaning, that co-commented $46\,198\,446$ times. Using this dataset we build an undirected network, where two users (nodes) are connected if and only if they commented under the same post at least once. The largest connected component of the co-commenting network has $G = 220\,275$ nodes and $E = 46\,193\,632$ links.

News: The dataset was built by considering a set of Facebook pages of news outlets listed by the Europe Media Monitor. By using the Facebook Graph API, all the posts and comments related to these pages in the period 2010-2015 were downloaded. Facebook pages are labelled according to the annotation obtained by MBFC. The dataset without annotations has been previously explored [3]. We consider $15\,540$ posts by 180 pages categorized from Left to Right (Left (12), Left-Center (80), Least-Biased (42), Right-Center (33), Right (13)). Such posts were co-commented $13\,525\,230$ times by 38663 active users (users with at least 3 likes and 3 comments), for which we identify the individual leaning. Using this dataset we build a undirected network, where two users (nodes) are connected if and only if they commented under the same post at least once. The largest connected component of the co-interaction network has $G = 38\,594$ nodes and $E = 13\,525\,119$ links.

5.5.4 Reddit

Politics: We consider $353\,864$ comments and submissions posted on the subreddit *politics* in the year 2017. From comments under submissions we reconstructed a directed network formed by $N = 240\,455$ users and $E = 5\,030\,565$ directed edges, where each edge represents a direct reply to a comment. The largest weakly connected component includes more than 99% of nodes. We exploited the classification retrieved from MBFC to identify the individual leaning of $N_c = 37\,148$ users, that is considered as a scalar feature of the node.

The Donald: We consider 1.234M comments and submissions posted on the subreddit *The_Donald* in the year 2017. From comments a submissions we reconstructed a directed network formed by $N = 138\,617$ users and $E = 5\,025\,290$ directed edges, where each edge represents a direct reply to a comment. The largest weakly connected component includes more than 99% of nodes. We exploited the classification retrieved from MBFC to identify the individual leaning of $N_c = 21\,905$ users.

News: We consider $723\,235$ comments and submissions posted on the subreddit *news* in the year 2017. From comments a submissions we reconstructed a directed network formed by $N = 179\,549$ users and $E = 1\,070\,589$ directed edges, where each edge represents a direct reply to a comment. The largest weakly connected component includes more than 99% of nodes. We exploited the classification retrieved from MBFC to identify the individual leaning of $N_c = 36\,875$ users.

Gab

The dataset, downloaded from `https://files.pushshift.io/gab`, spans from the first Gab post (occurred in 2016) to the late 2018 and it includes data regarding post-reply relationships, number of upvotes of posts, repost or replies and their timestamps. We selected all the contents (post, reply, quote) in the time window ranging from 11/2017 to 10/2018, that is $C = 13\,580\,937$ unique pieces of content created by $N = 165\,162$ unique users. We consider all the post that have a link to an external source, for an amount of $3\,302\,621$ posts (excluding YouTube links). By extracting the domain from each link we obtain a set of $75\,436$ unique domains. In this set, 1650 unique domains for a total of $1\,454\,502$ URLs (44%) were labelled using the classification provided by MBFC. We identified the in-

Table 5.1: For each data set, we report: the starting date of collection T_0, time span T expressed in days (d) or years (y), number of unique contents C, number of users N, coverage n_c (fraction of users with classified leaning), size of the giant component G and average node degree $\langle k \rangle$.

Media	Data set	T_0	T	C	N	n_c	G	$\langle k \rangle$
	Gun control	06/2016	14 d	19 M	3963	0.93	3717	798
Twitter	Obamacare	06/2016	7 d	39 M	8703	0.90	8703	1405
	Abortion	06/2016	7 d	34 M	7401	0.95	6828	478
	Sci/Cons	01/2010	5 y	75 172	183 378	1.00	181960	228
Facebook	Vaccines	01/2010	7 y	94 776	221 758	1.00	220275	419
	News	01/2010	6 y	15 540	38 663	1.00	38594	700
	Politics	01/2017	1 y	353 864	240 455	0.15	240455	9
Reddit	The Donald	01/2017	1 y	1.234 M	138 617	0.16	138617	31
	News	01/2017	1 y	723 235	179 549	0.20	179549	3
Gab	Gab	11/2017	1 y	13 M	165 162	0.13	20701	328

dividual leaning of $N_c = 31\,286$ users. We also reconstructed the interaction network using co-commenting as a proxy, that is, two users are connected if and only if they commented under the same post at least once. The largest connected component of the network includes $G = 20\,701$ nodes, about 66% of the users with assigned leaning, and $E = 8\,273\,412$ edges. The individual leaning x_i is considered as a scalar feature of the node.

5.5.5 Analysis for other Datasets

In this section we report the results obtained for other four data sets not discussed in the sections before for the sake of brevity, namely “Science and Conspiracy” (Facebook), “Gun control” (Twitter), “Obamacare” (Twitter) and ‘The Donald” (Reddit). The techniques and the pipeline is the same used for the datasets analyzed in the previous sections.

Science and Conspiracy

Figure 5.7 displays the results obtained for the Facebook dataset called “Science and Conspiracy”, described in Section 5.5.4. Panel (a) shows the joint distribution of the leaning of users, x, against the average leaning of their neighborhood X^N. We note that the community referred to as “Science”, to which is associated a leaning of -1, is much smaller than the community called ”Conspiracy” and for this reason it is not clearly visible in the density plot but only in the histograms at its margins. Panel (b) shows the size and average leaning of communities detected by the Louvain algorithm.

Panels (c) and (d) show the results of the SIR dynamics: the average leaning $\langle\mu(x)\rangle$ of the influence sets reached by users with leaning x, for two different values of the infection probability, while the recovery rate is fixed $\nu = 0.2$. Size and color of each point is related to the average size of the influence sets.

Guncontrol

Figure 5.8 shows the results obtained for the Twitter dataset “Gun control”, described in Section 5.5.4. Panel (a) shows the joint distribution of the leaning of users, x, against the average leaning of their neighborhood X^N, in which two different regions are clearly visible. Panel (b) shows the size and average leaning of communities detected by the Louvain algorithm.

Panels (c) and (d) show the results of the SIR dynamics: the average leaning $\langle\mu(x)\rangle$ of the influence sets reached by users with leaning x, for two different values of the infection probability, while the recovery rate is fixed $\nu = 0.2$. Size and color of each point is related to the average size of the influence sets.

Obamacare

Figure 5.9 shows the results obtained for the Twitter dataset referred to as "Obamacare", described in Section 5.5.4. Panel (a) shows the joint distribution of the leaning of users, x, against the average leaning of their neighborhood X^N, in which two interconnected regions are clearly visible. Panel (b) shows the size and average leaning of communities detected by the Louvain algorithm.

Panels (c) and (d) show the results of the SIR dynamics: the average leaning $\langle\mu(x)\rangle$ of the influence sets reached by users with leaning x, for two different values of the infection probability, while the recovery rate is fixed $\nu = 0.2$. Size and color of each point is related to the average size of the influence sets.

TheDonald

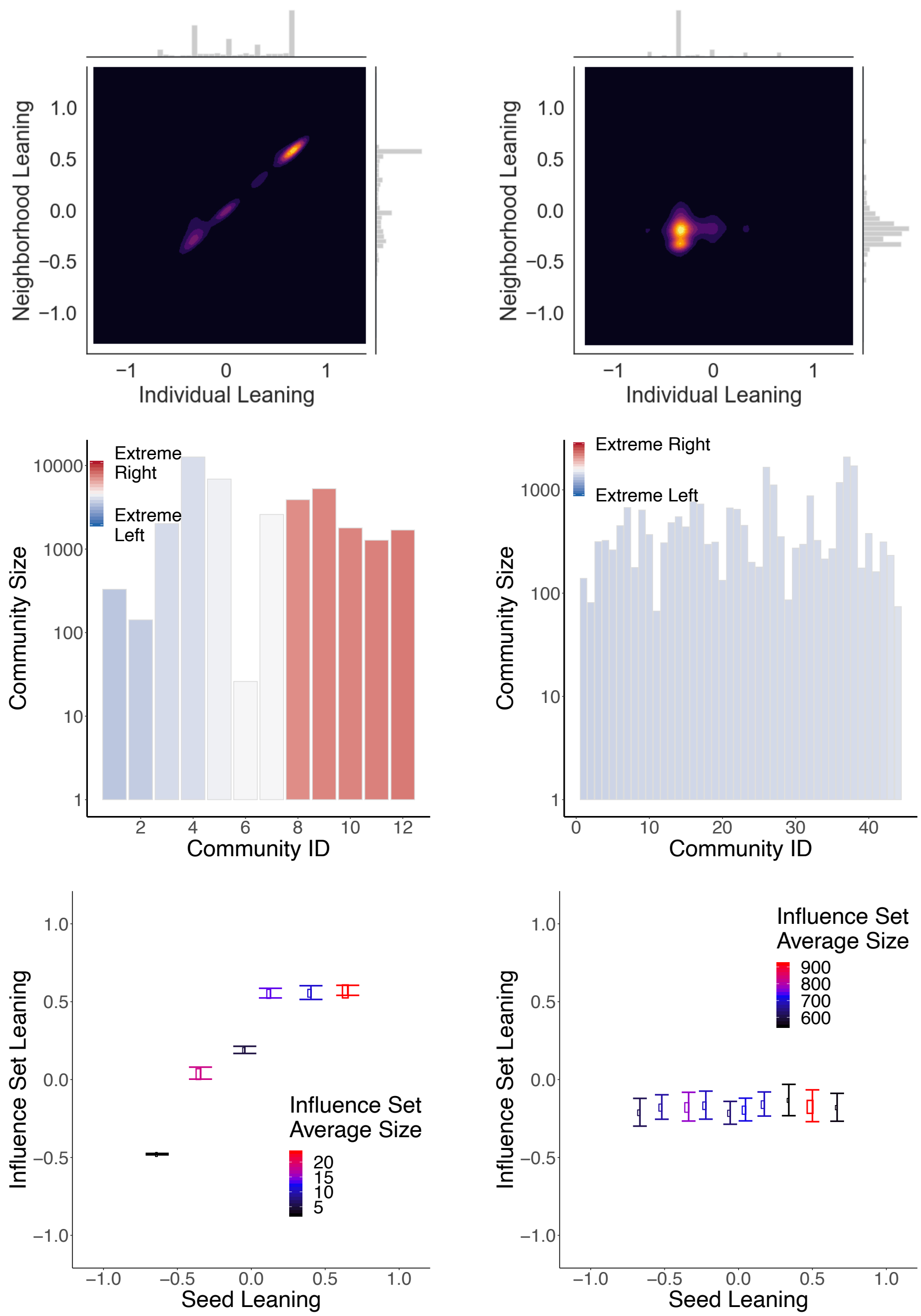
Figure 5.10 shows the results obtained for the Reddit dataset “The Donald”, described in Section 5.5.4. Panel (a) displays the joint distribution of the leaning of users, x, against the average leaning of their neighborhood X^N, showing a unique region spanning most of the x-axis and concentrated on the values around 0.25 on the y-axis. Such a region is also characterized by few peaks of leaning (spanning mainly from Center to Extreme Right) that are displayed in the histogram on the top margin. Panel (b) shows the size and average leaning of communities detected by the Louvain algorithm.

Panels (c) and (d) show the results of the SIR dynamics: the average leaning $\langle\mu(x)\rangle$ of the influence sets reached by users with leaning x, for two different values of the infection probability, while the recovery rate is fixed $\nu = 0.2$. Size and color of each point is related to the average size of the influence sets.

5.5.6 Robustness of the SIR Dynamics

In this section, we provide additional results for the SIR dynamics run with different parameters on the 6 data sets considered in section 5.3, namely “Abortion” on Twitter, “Politics” and “News” on Reddit, “Vaccines” and “News” on Facebook, and Gab.

The results, reported in Fig. 5.11, are qualitatively identical to the ones in the main paper and are reported here for the sake of brevity. Details about the parameters used in the simulations are provided in the caption of Fig. 5.11.



(a) Facebook **(b)** Reddit

Fig. 5.4: Direct comparison of news consumption on Facebook (left column) and Reddit (right column) Joint distribution of the leaning of users x and the average leaning of their nearest-neighbor x^N (top row), size and average leaning of communities detected in the interaction networks (middle row), and average leaning $\langle \mu(x) \rangle$ of the influence sets reached by users with leaning x, by running SIR dynamics (bottom row) with parameters $\beta = 0.05\langle k \rangle$ for panel (a) and $\beta = 0.006\langle k \rangle$ for panel (b) and $\nu = 0.2$ for both. Facebook presents a highly segregated structure w.r.t. Reddit

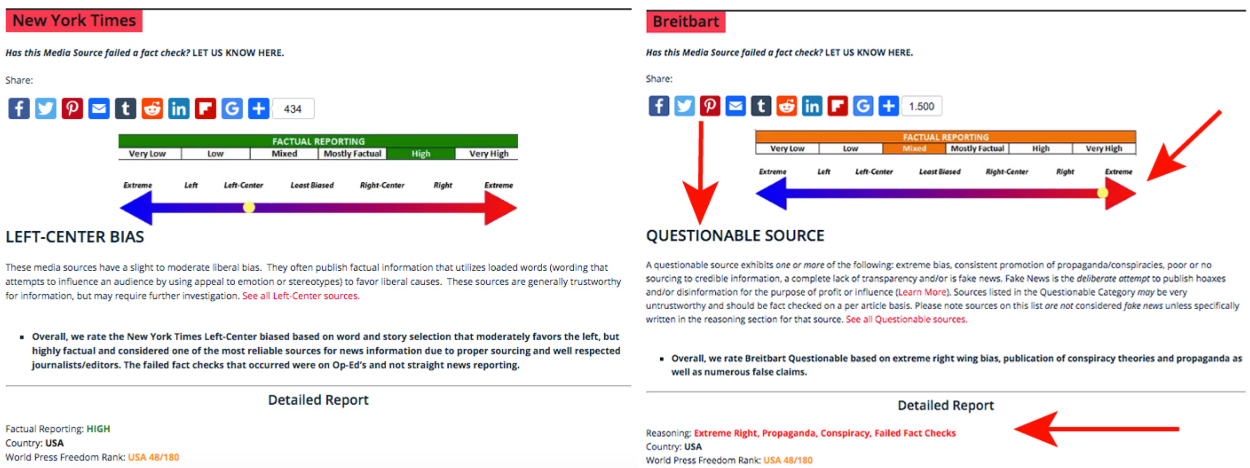


Fig. 5.5: Example of the web page of MBFC for two news outlets, namely New York Time and Breitbart. Notice that, although Breitbart is labeled as ”Questionable”, an explicit leaning appears in its description.

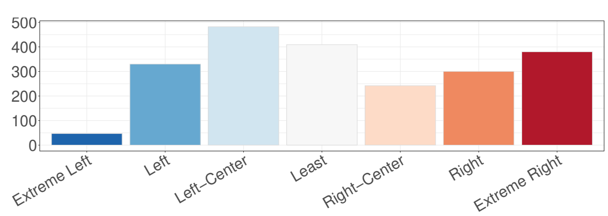


Fig. 5.6: Distribution of the leanings assigned to each source, ranging from Extreme Left (numerical value: -1, colored in blue) to Extreme Right (numerical value: +1, colored in red).

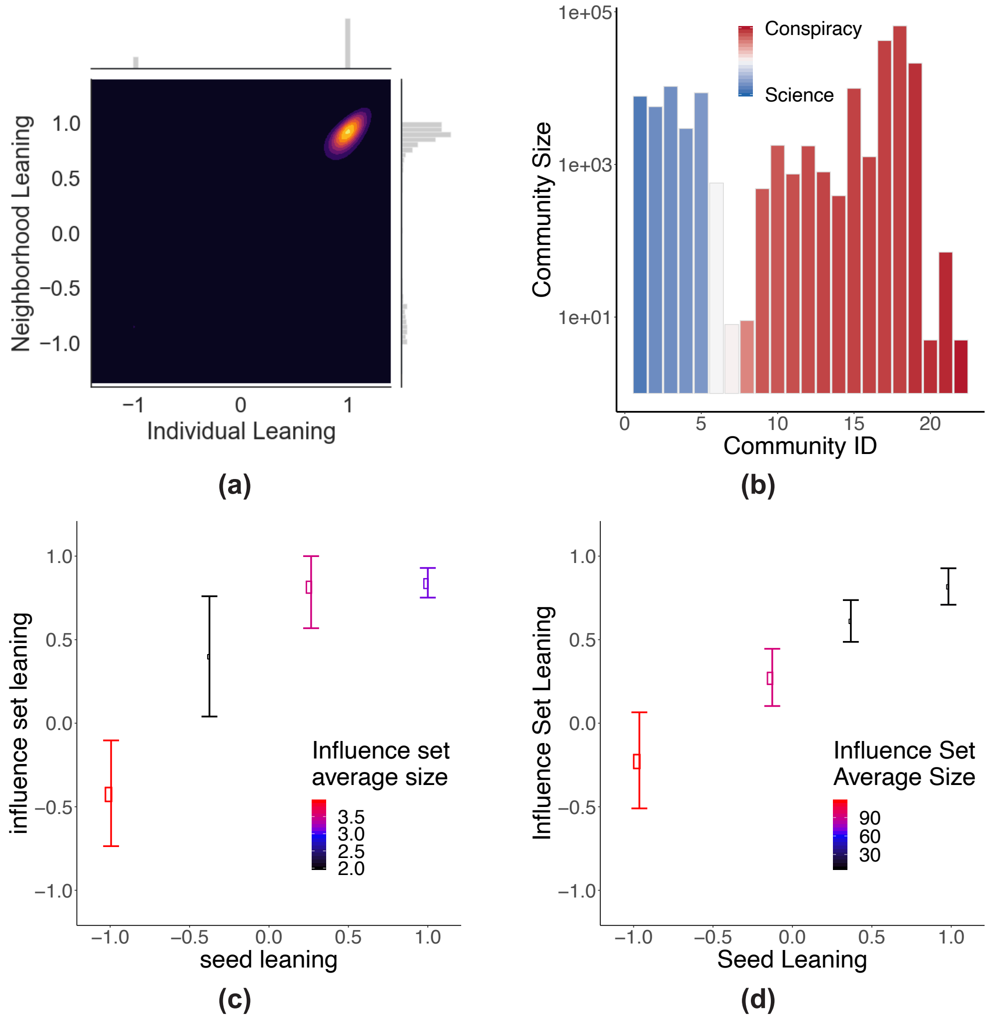


Fig. 5.7: Science vs Conspiracy. Panel (a): Individual leaning versus neighborhood leaning. Panel (b): Community detection. Panel (c) and (d): average leaning $\langle \mu(x) \rangle$ of the influence sets reached by users with leaning x, for infection probability $\beta = 0.01\langle k \rangle^{-1}$ and $\beta = 0.02\langle k \rangle^{-1}$, respectively, where $\langle k \rangle$ is the average degree of the network.

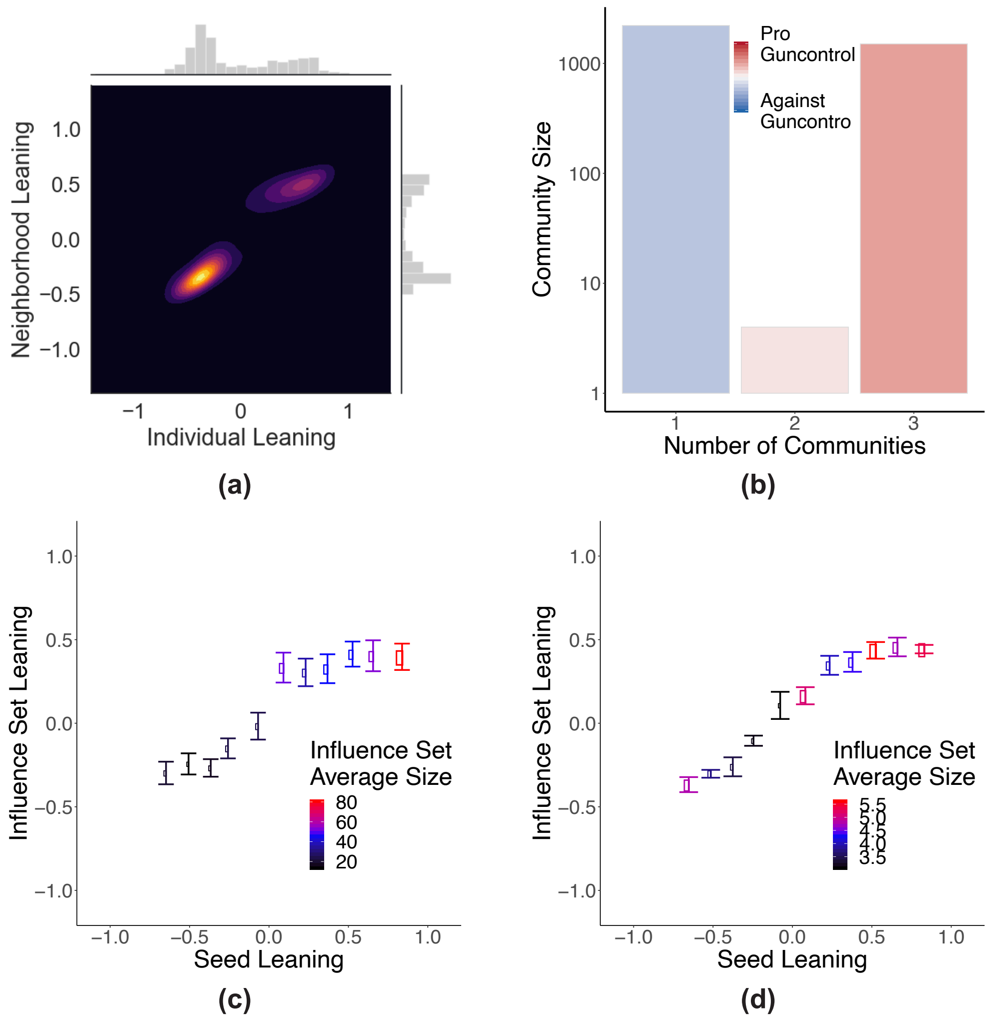


Fig. 5.8: Gun control. Panel (a): Individual leaning versus neighborhood leaning. Panel (b): Community detection. Panel (c) and (d): average leaning $\langle \mu(x) \rangle$ of the influence sets reached by users with leaning x, for infection probability $\beta = 0.1\langle k \rangle^{-1}$ and $\beta = 0.2\langle k \rangle^{-1}$, respectively, where $\langle k \rangle$ is the average degree of the network.

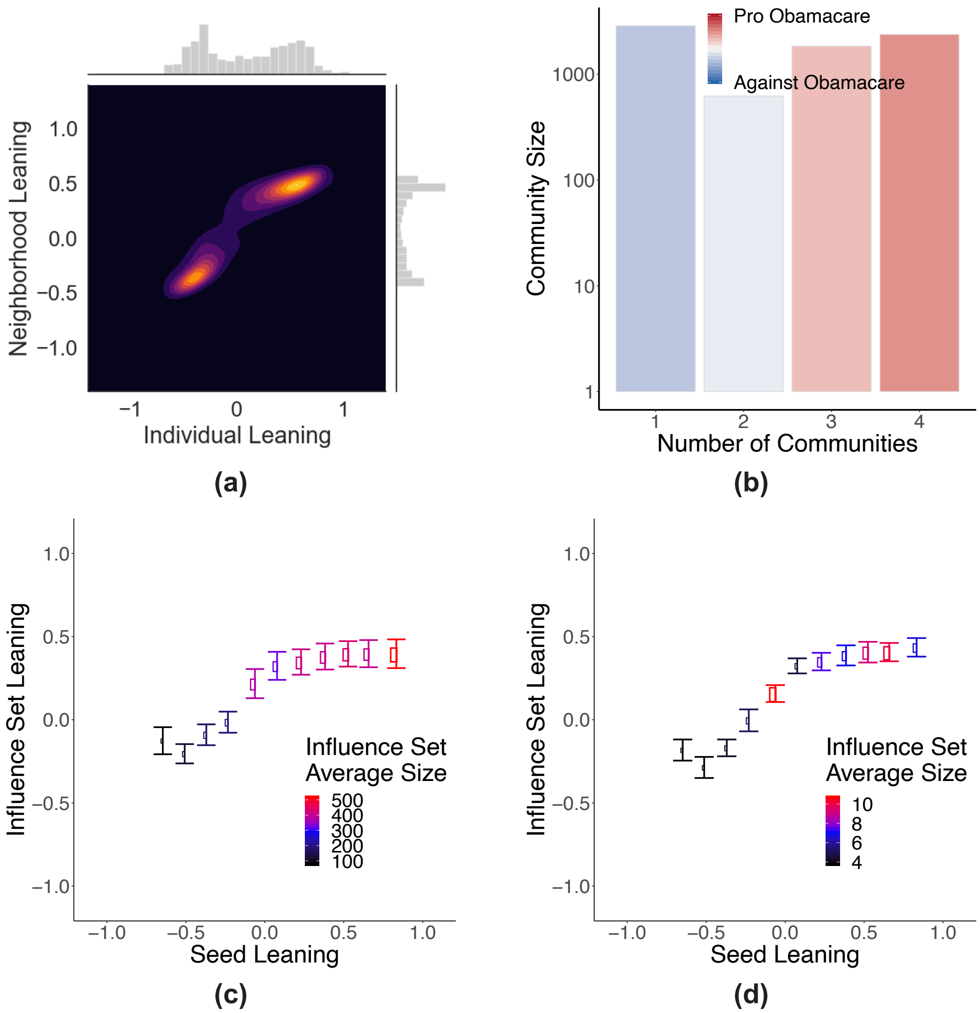


Fig. 5.9: Obamacare. Panel (a): Individual leaning versus neighborhood leaning. Panel (b): Community detection. Panel (c) and (d): average leaning $\langle\mu(x)\rangle$ of the influence sets reached by users with leaning x, for infection probability $\beta = 0.1\langle k\rangle^{-1}$ and $\beta = 0.2\langle k\rangle^{-1}$, respectively, where $\langle k\rangle$ is the average degree of the network.

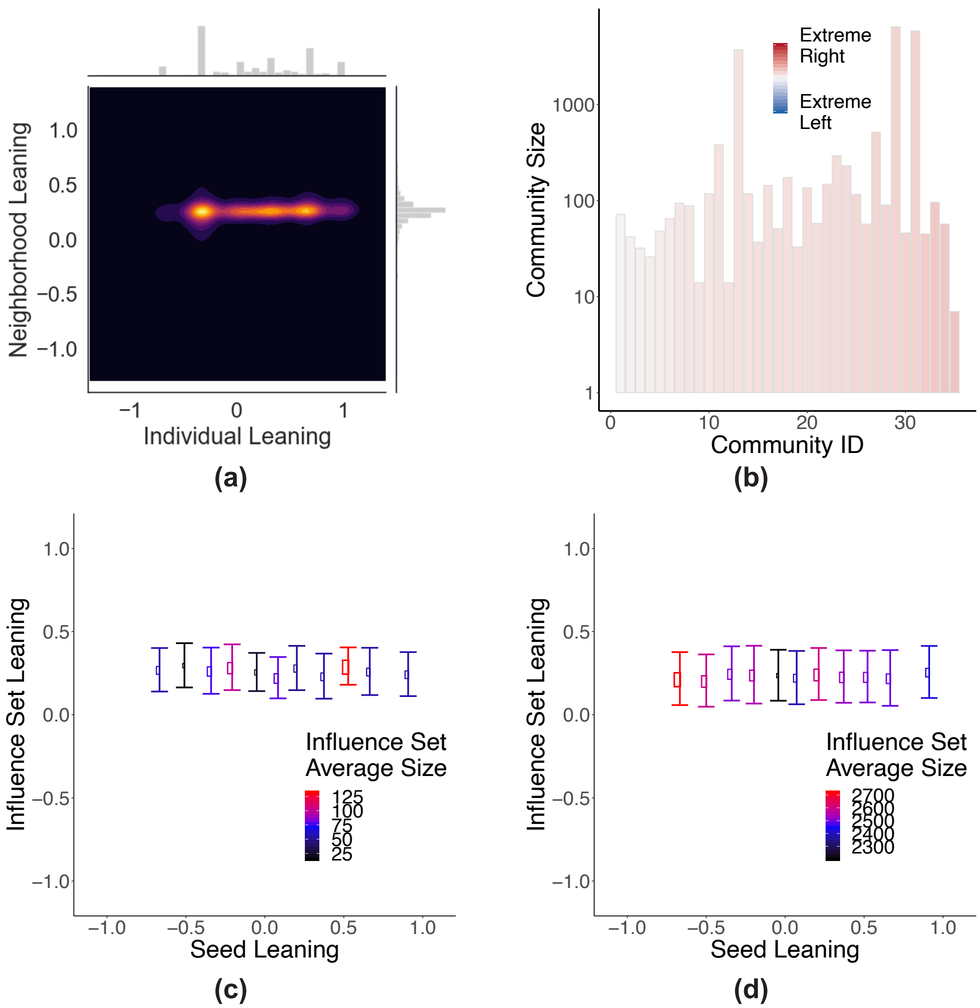


Fig. 5.10: The Donald. Panel (a): Individual leaning versus neighborhood leaning. Panel (b): Community detection. Panel (c) and (d): average leaning $\langle\mu(x)\rangle$ of the influence sets reached by users with leaning x, for infection probability $\beta = 0.0067\langle k\rangle^{-1}$ and $\beta = 0.013\langle k\rangle^{-1}$, respectively, where $\langle k\rangle$ is the average degree of the network.

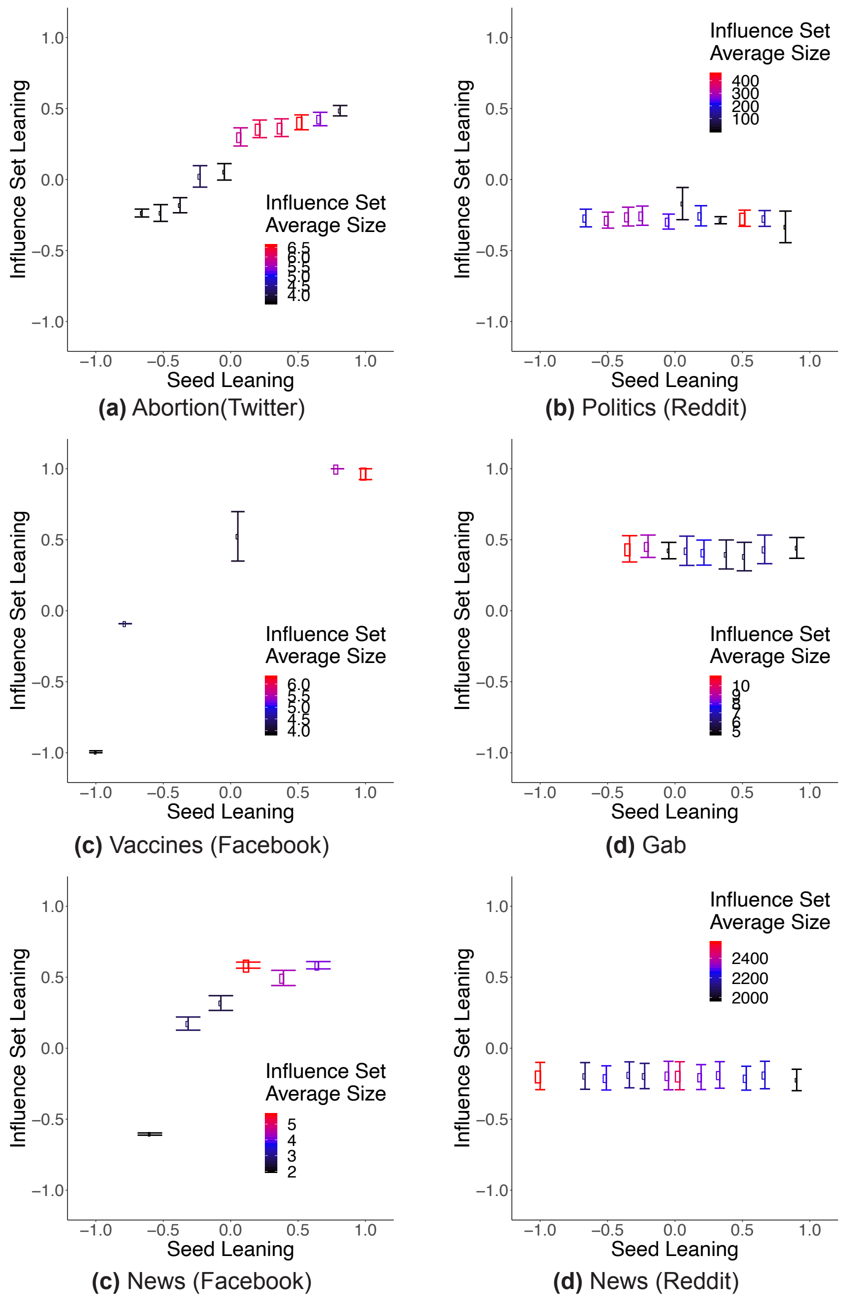


Fig. 5.11: Additional results of the SIR dynamics for the six data sets considered in the main paper. Average leaning $\langle \mu(x) \rangle$ of the influence sets reached by users with leaning x, for infection probability $\beta = 0.05\langle k \rangle^{-1}$ (Abortion on Twitter, panel (a)), $\beta = 0.005\langle k \rangle^{-1}$ (Politics on Reddit, panel (b)), $\beta = 0.02\langle k \rangle^{-1}$ (Vaccines on Facebook, panel (c)), $\beta = 0.025\langle k \rangle^{-1}$ (Gab, panel (d)), $\beta = 0.025\langle k \rangle^{-1}$ (News on Facebook, panel (e)), $\beta = 0.01\langle k \rangle^{-1}$ (News on Reddit, panel (f)), while the recovery rate is fixed $\nu = 0.2$. Size and color of each point is related to the average size of the influence sets.

CHAPTER 6

MEASURING THE RISE OF COVID-19 DEBATE ON SOCIAL MEDIA

The previous chapter was dedicated to the study of the echo chambers effect on different platform and around several debated topics. Our study showed that users tend to interact with peers holding the same beliefs and thus cluster together into ideologically homogeneous groups. This tendency may be fostered by platform feed algorithms that strongly suggest contents adhering to users beliefs. Moreover, our analysis also revealed important insights on the spreading patterns of news inside echo chambers and how each platform may experience different diffusion patterns. Hence, one question of primary interest is how events of high public interest are debated on different platforms.

In this chapter, we analyze the impact of a dramatic event such as the COVID-19 outbreak on different online environments. We compare the information consumption about COVID-19 in five different social platforms considering both the contents created and the engagement they receive. We also study the diffusion of news from reliable and questionable sources and compare their popularity across platforms.

6.1 Introduction

The World Health Organization (WHO) defined the SARS-CoV-2 virus outbreak as a severe global threat[147]. As foreseen in 2017 by the global risk report of the World Economic forum, global risks are interconnected. In particular, the case of the COVID-19 epidemic (the infectious disease caused by the most recently discovered human coronavirus) is showing the critical role of information diffusion in a disintermediated news cycle [1].

The term *infodemic* [24], [148] has been coined to outline the perils of misinformation phenomena during the management of disease outbreaks [149]–[151], since it could even speed up the epidemic process by influencing and fragmenting social response [152]. As an example, CNN has recently anticipated a rumor about the possible lock-down of Lombardy (a region in northern Italy) to prevent pandemics[153], publishing the news hours before the official communication from the Italian Prime Minister. As a result, people overcrowded trains and airports to escape from Lombardy toward the southern regions before the lock-down was put in place, disrupting the government initiative aimed to contain the epidemics and potentially increasing contagion. Thus, an important research challenge is to determine how people seek or avoid information and how those decisions affect their behavior [123], particularly when the news cycle – dominated by the disintermediated diffusion of information – alters the way information is consumed and reported on.

The case of the COVID-19 epidemic shows the critical impact of this new information environment. The information spreading can strongly influence people's behavior and alter the effectiveness of the countermeasures deployed by governments. To this respect, models to forecast virus spreading are starting to account for the behavioral response of the population with respect to public health interventions and the communication dynamics behind content consumption [152], [154], [155].

Social media platforms such as YouTube and Twitter provide direct access to an unprecedented amount of content and may amplify rumors and questionable information. Taking into account users' preferences and attitudes, algorithms mediate and facilitate content promotion and thus information spreading [156]. This shift from the traditional news paradigm profoundly impacts the construction of social perceptions [3] and the framing of narratives; it influences policy-making, political communication, as well as the evolution of public debate [145], [157], especially when issues are controversial [2]. Users online tend to acquire information adhering to their worldviews [4], [40], to ignore dissenting information [124], [158] and to form polarized groups around shared narratives [57], [109]. Furthermore, when polarization is high, misinformation might easily proliferate

[68], [159]. Some studies pointed out that fake news and inaccurate information may spread faster and wider than fact-based news [14]. However, this might be platform-specific effect. The definition of “Fake News” may indeed be inadequate since political debate often resorts to labelling opposite news as unreliable or fake [16]. Studying the effect of the social media environment on the perception of polarizing topics is being addressed also in the case of COVID-19. The issues related to the current infodemics are indeed being tackled by the scientific literature from multiple perspectives including the dynamics of hatespeech and conspiracy theories [160], [161], the effect of bots and automated accounts [162], and the threats of misinformation in terms of diffusion and opinions formation [163], [164].

In this work we provide an in-depth analysis of the social dynamics in a time window where narratives and moods in social media related to the COVID-19 have emerged and spread. While most of the studies on misinformation diffusion focus on a single platform [2], [14], [18], the dynamics behind information consumption might be particular to the environment in which they spread on. Consequently, in this study we perform a comparative analysis on five social media platforms (Twitter, Instagram, YouTube, Reddit and Gab) during the COVID-19 outbreak. The dataset includes more than 8 million comments and posts over a time span of 45 days. We analyze user engagement and interest about the COVID-19 topic, providing an assessment of the discourse evolution over time on a global scale for each platform. Furthermore, we model the spread of information with epidemic models, characterizing for each platform its basic reproduction number (R_0), i.e. the average number of secondary cases (users that start posting about COVID-19) an "infectious" individual (an individual already posting on COVID-19) will create. In epidemiology, $R_0 = 1$ is a threshold parameter. When $R_0 < 1$ the disease will die out in a finite period of time, while the disease will spread for $R_0 > 1$. In social media, $R_0 > 1$ will indicate the possibility of an infodemic.

Finally, coherently with the classification provided by the fact-checking organization Media Bias/Fact Check [165] that classifies news sources based on the truthfulness and bias of the information published, we split news outlets into two groups. These groups are either associated to the diffusion of (mostly) reliable or (mostly) questionable contents and we characterize the spreading of information regarding COVID-19 relying on this classification. We find that users in mainstream platforms are less susceptible to the diffusion of information from questionable sources and that information deriving from news outlets marked either as reliable or questionable do not present significant difference in the way it spreads.

Our findings suggest that the interaction patterns of each social media combined with the peculiarity of the audience of each platform play a pivotal role in information and misinformation spreading. We conclude the paper by measuring rumor's amplification parameters for COVID-19 on each social media platform.

6.2 Results

We analyze mainstream platforms such as Twitter, Instagram and YouTube as well as less regulated social media platforms such as Gab and Reddit. Gab is a crowd-funded social media whose structure and features are Twitter-inspired. It performs very little control on content posted; in the political spectrum, its user base is considered to be far-right. Reddit is an American social news aggregation, web content rating, and discussion website based on collective filtering of information.

We perform a comparative analysis of information spreading dynamics around the same argument in different environments having different interaction settings and audiences. We collect all pieces of content related to COVID-19 from the 1st of January to the 14th of February. Data have been collected filtering contents accordingly to a selected sample of Google Trends' COVID-19 related queries such as: *coronavirus*, *coronavirusoutbreak*, *imnotavirus*, *ncov*, *ncov*-19, *pandemic*, *wuhan*. The deriving dataset is then composed of 1,342,103 posts and 7,465,721 comments produced by 3,734,815 users. For more details regarding the data collection refer to section 6.4.

6.2.1 Interaction Patterns

First, we analyze the interactions (i.e., the engagement) that users have with COVID-19 topics on each platform. The upper panel of Figure 6.1 shows users' engagement around the COVID-19 topic. Despite the differences among platforms, we observe that they all display a rather similar distribution of the users' activity characterized by a long tail. This entails that users behave similarly for what concern the dynamics of reactions and content consumption. Indeed, users' interactions with the COVID-19 content present attention patterns similar to any other topic [166]. The highest volume of interactions in terms of posting and commenting can be observed on mainstream platforms such as YouTube and Twitter.

Then, to provide an overview of the debate concerning the disease outbreak, we extract and analyze the topics related to the COVID-19 content by means of Natural Language Processing techniques. We build word embedding for the text corpus of each platform, i.e. a word vector representation in which words sharing common contexts are in close proximity. Moreover, by running clustering procedures on these vector representations, we separate groups of words and topics that are perceived as more relevant for the COVID-19 debate. For further details refer to section 6.4. The results (Figure 6.1, middle panel) show that topics are quite similar across each social media platform. Debates range from comparisons

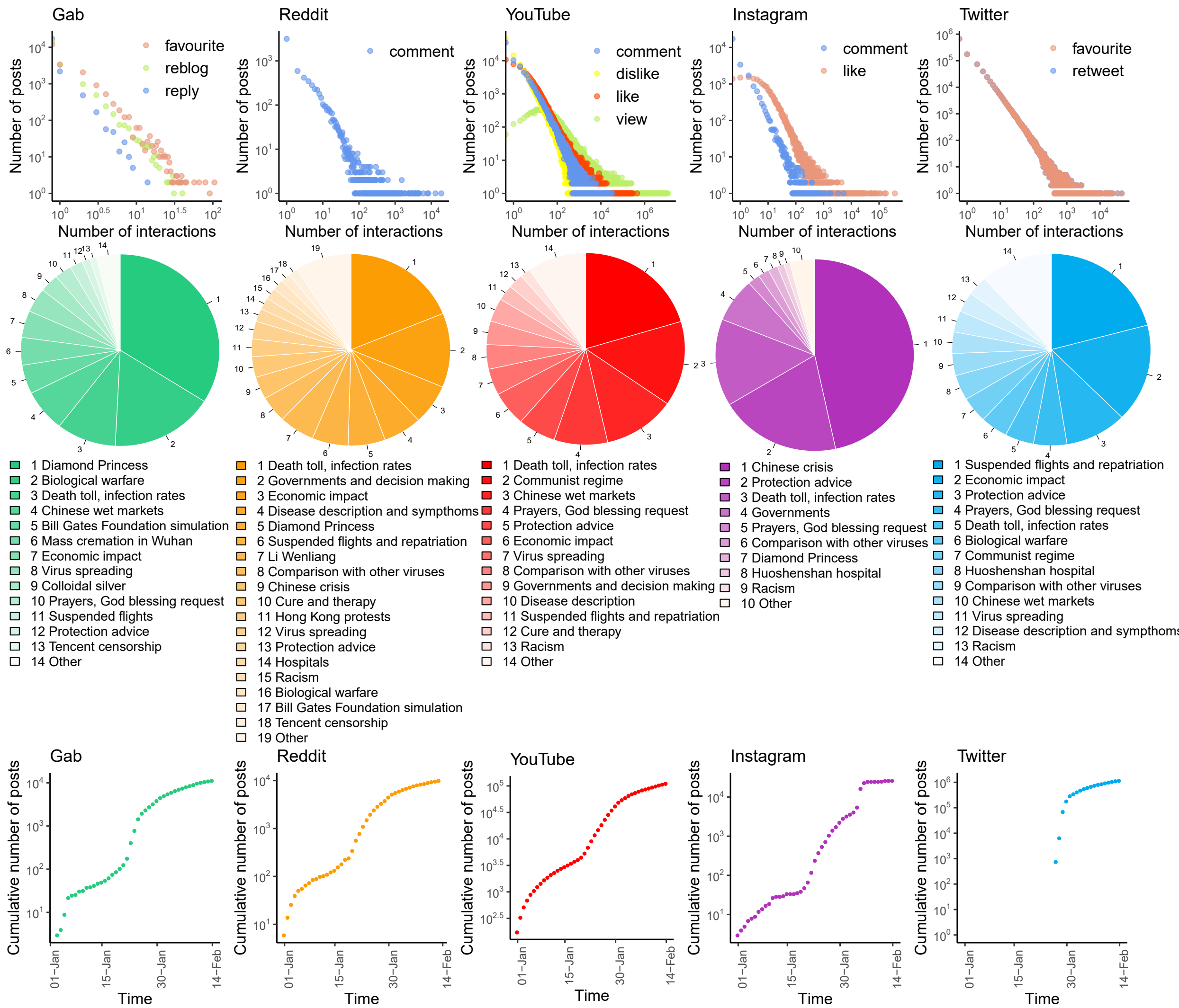


Fig. 6.1: Upper panel: activity (likes, comments, reposts, etc..) distribution for each social media. Middle panel: most discussed topics about COVID-19 on each social media. Lower panel: cumulative number of content (posts, tweets, videos, etc..) produced from the 1^{st} of January to the $14^t h$ of February. Due to the Twitter API limitations in gathering past data, the first data point for Twitter is dated January 27^{th}.

to other viruses, requests for God blessing, up to racism, while the largest volume of interaction is related to the lock-down of flights.

Finally, to characterize user engagement with the COVID-19 on the five platforms, we compute the cumulative number of new posts each day (Figure 6.1, lower panel). For all platforms, we find a change of behavior around the 20^{th} of January, that is the day that the World Health Organization (WHO) issued its first situation report on the COVID-19 [167]. The largest increase in the number of posts is on the 21^{st} of January for Gab, the 24^{th} January for Reddit, the 30^{th} January for Twitter, the 31^{th} January for YouTube and the 5^{th} of February for Instagram. Thus, social media platforms seem to have specific timings for content consumption; such patterns may depend upon the difference in terms of audience and interaction mechanisms (both social and algorithmic) among platforms.

6.2.2 Information Spreading

Efforts to simulate the spreading of information on social media by reproducing real data have mostly applied variants of standard epidemic models [168]–[171]. Coherently, we analyze the observed monotonic increasing trend in the way new users interact with information related to the COVID-19 by using epidemic models. Unlike previous works, we do not only focus on models that imply specific growth mechanisms, but also on phenomenological models that emphasize the reproducibility of empirical data [172].

Most of the epidemiological models focus on the basic reproduction number R_0, representing the expected number of new infectors directly generated by an

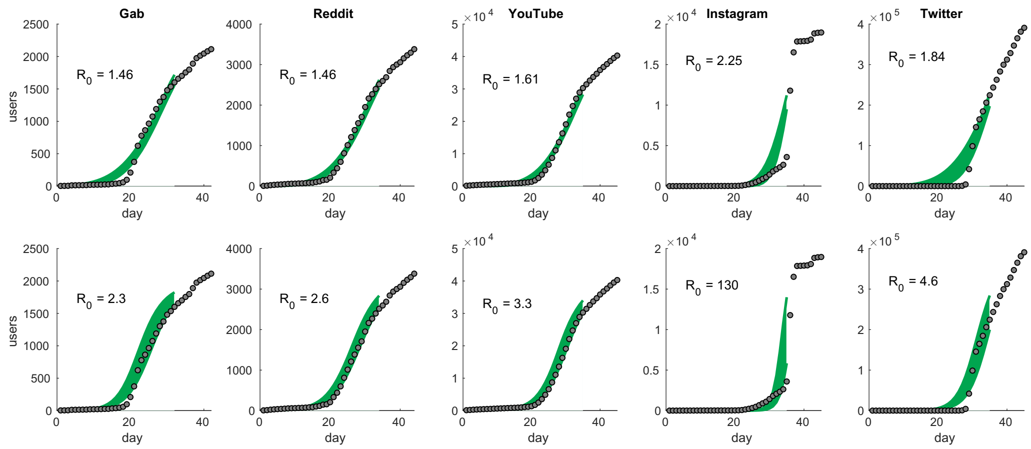


Fig. 6.2: Growth of the number of authors vs time. Time is expressed in number of days since 1^{st} Jan 2020 (day 1). Shaded areas represents $[5\%, 95\%]$ estimates of the models obtained via bootstrapping least square estimates of the EXP model (upper panels) and of the SIR model (lower panels). For details the SIR and the EXP model, see SI.

infected individual for a given time period [55]. An epidemic occurs if $R_0 > 1$, – i.e., if an exponential growth in the number of infections is expected at least in the initial phase. In our case, we try to model the growth in number of people publishing a post on a subject as an infective process, where people can start publishing after being exposed to the topic. While in real epidemics $R_0 > 1$ highlights the possibility of a pandemic, in our approach $R_0 > 1$ indicates the emergence of an infodemic. We model the dynamics both with the phenomenological model of [53] (from now on referred to as the EXP model) and with the standard SIR (Susceptible, Infected, Recovered) compartmental model [54]. Further details on the modeling approach can be found in section 6.4.

As shown in Figure 6.2, each platform has its own basic reproduction num-

	Gab	Reddit	YouTube	Instagram	Twitter
$\mathbf{R_0^{EXP}}$	$[1.42, 1.52]$	$[1.44, 1.51]$	$[1.56, 1.70]$	$[2.02, 2.64]$	$[1.65, 2.06]$
$\mathbf{R_0^{SIR}}$	$[2.2, 2.5]$	$[2.4, 2.8]$	$[3.2, 3.5]$	$[1.1x10^2, 1.6x10^2]$	$[4.0, 5.1]$

Table 6.1: $[5\%, 95\%]$ interval of confidence R_0 as estimated from bootstrapping the least square fits parameter of the EXP and of the SIR model. Notice that, due to the steepness of the growth of the number of new authors in Instagram, R_0 assumes unrealistic values $\sim 10^2$ for the SIR model.

ber R_0. As expected, all the values of R_0 are supercritical - even considering confidence intervals (Table 6.1) - signaling the possibility of an infodemic. This observation may facilitate the prediction task of information spreading during critical events. Indeed, according to this result we can consider information spreading patterns on each social media to predict social response when implementing crisis management plans.

While R_0 is a good proxy for the engagement rate and a good predictor for epidemic-like information spreading, social contagion phenomena might be in general more complex [133], [173], [174]. For instance, in the case of Instagram, we observe an abrupt jump in the number of new users that cannot be explained with continuous models like the standard epidemic ones; accordingly, the SIR model estimates a value of $R_0 \sim 10^2$ that is way beyond what has been observed in any real-world epidemic.

6.2.3 Questionable VS Reliable Information Sources

We conclude our analysis by comparing the diffusion of information from questionable and reliable sources on each platform. We tag links as reliable or questionable according to the data reported by the independent fact-checking organization Media Bias/Fact Check [165]. In order to clarify the limits of an approach that is based on labelling news outlets rather than single articles, as for instance performed in [18], [19], we report the definitions used in this paper for questionable and reliable information sources. In accordance with the criteria established by MBFC, by questionable information source we mean a news outlet systematically showing one or more of the following characteristics: extreme bias, consistent promotion of propaganda/conspiracies, poor or no sourcing to credible information, information not supported by evidence or unverifiable, a complete lack of transparency and/or fake news. By reliable information sources we mean news outlets that do not show any of the aforementioned characteristics. Such outlets can anyway produce contents potentially displaying a bias towards liberal/conservative opinion, but this does not compromise the overall reliability of the source.

Figure 6.3 shows, for each platform, the plots of the cumulative number of posts and reactions related to reliable sources versus the cumulative number of posts and interactions referring to questionable sources. By interactions we mean

	$\mathcal{E}^U$	$\mathcal{E}^R$	α
Gab	5.6	1.4	3.9
Reddit	22.7	40.1	0.55
Twitter	15.1	15.6	0.97
YouTube	1.4×10^4	3.9×10^4	0.35

Table 6.2: The average engagement of a post is the number of reactions expected for a post and is a measure of how much a post is amplified in each social media platform. The average engagement $\mathcal{E}^U$ (for unreliable post) and $\mathcal{E}^R$ (for reliable post) vary from platform to platform, and are the largest in Twitter and the lowest in Gab. The coefficient of relative amplification $\alpha = \mathcal{E}^U/\mathcal{E}^R$ measures whether a social media amplifies more unreliable ($\alpha > 1$) or reliable ($\alpha < 1$) posts. Among more popular social media platforms, we notice that Twitter is the most neutral ($\alpha \sim 1\%$ i.e. $\mathcal{E}^U \sim \mathcal{E}^R$), while YouTube amplifies unreliable sources less ($\alpha \sim 4/10$). Among less popular social media platforms, Reddit reduces the impact of unreliable sources ($\alpha \sim 1/2$) while Gab strongly amplifies them ($\alpha \sim 4$).

the overall reactions, e.g. likes or other form or endorsement and comments, that can be performed with respect to a post on a social platform. Surprisingly, all the posts show a strong linear correlation, i.e., the number of posts/reactions relying on questionable and reliable sources grows with the same pace inside the same social media platform. We observe the same phenomenon also for the engagement with reliable and questionable sources. Hence, the growth dynamics of posts/interactions related to questionable news outlets is just a re-scaled version of the growth dynamics of posts/reactions related to reliable news outlets; however, the re-scaling factor ρ (i.e., the fraction of questionable over reliable) is strongly dependent on the platform.

In particular, we observe that in mainstream social media the number of posts produced by questionable sources represents a small fraction of posts produced

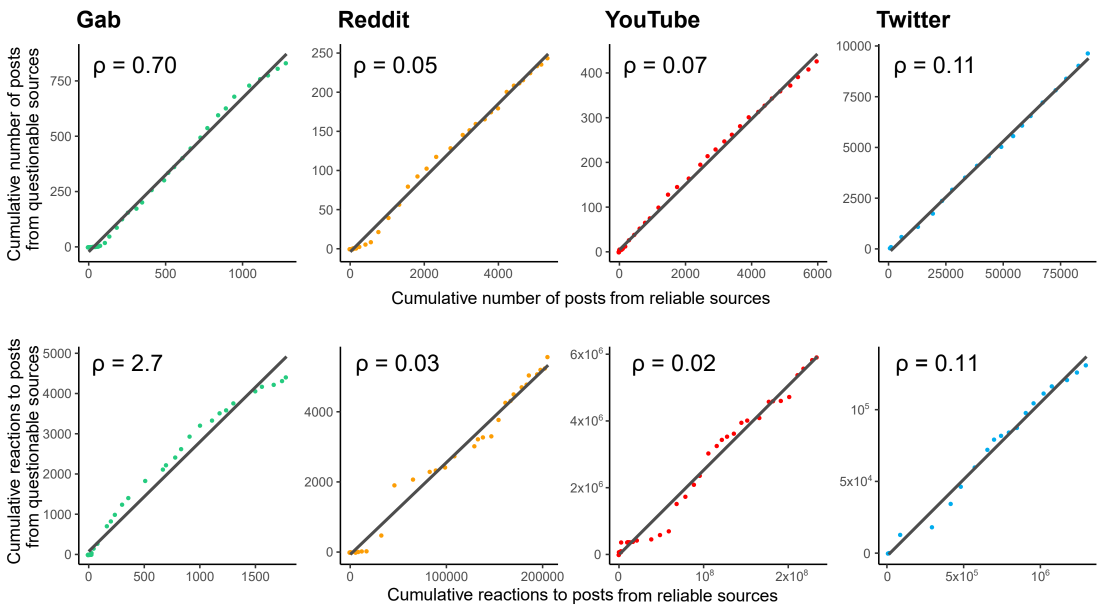


Fig. 6.3: Upper panels: plot of the cumulative number of posts referring to questionable sources versus the cumulative number of posts referring to reliable sources. Lower panel: plot of the cumulative number of engagements relatives to questionable sources versus the cumulative number of engagements relative to reliable sources. Notice that a linear behavior indicates that the time evolution of questionable posts/engagements is just a re-scaled version of the time evolution of reliable posts/engagements. Each plot indicates the regression coefficients ρ, representing the ratio among the volumes of questionable and reliable posts (ρ^{post}) and engagements (ρ^{eng}). In more popular social media, the number of questionable posts represents a small fraction of the reliable ones; same thing happens in Reddit. Among less popular social media, a peculiar effect is observed in Gab: while the volume of questionable posts is just the $\sim 70\%$ of the volume of reliable ones, the volume of engagements for questionable posts is ~ 3 times bigger than the volume for reliable ones. Further details concerning the regression coefficients are reported in Methods.

by reliable ones; the same thing happens in Reddit. Among less regulated social media, a peculiar effect is observed in Gab: while the volume of posts from questionable sources is just the $\sim 70\%$ of the volume of posts from reliable ones, the volume of reactions for the former ones is ~ 3 times bigger than the volume for the latter ones. Such results hint the possibility that different platform react differently to information produced by reliable and questionable news outlets.

To further investigate this issue, we define the amplification factor $\mathcal{E}$ as the average number of reactions to a post; hence, $\mathcal{E}$ is a measure that quantifies the extent to which a post is amplified in a social media. We observe that the amplification $\mathcal{E}^U$ (for unreliable posts posts produced by questionable outlets) and $\mathcal{E}^R$ (for reliable posts posts produced by reliable outlets) vary from social media platform to social media platform and that assumes the largest values in YouTube and the lowest in Gab. To measure the permeability of a platform to posts from questionable/reliable news outlets, we then define the coefficient of relative amplification $\alpha = \mathcal{E}^U/\mathcal{E}^R$. It is a measure of whether a social media amplifies questionable ($\alpha > 1$) or reliable ($\alpha < 1$) posts. Results are shown in Table 6.2. Among mainstream social media, we notice that Twitter is the most neutral ($\alpha \sim 1$ i.e. $\mathcal{E}^U \sim \mathcal{E}^R$), while YouTube amplifies questionable sources less ($\alpha \sim 4/10$). Among less popular social media, Reddit reduces the impact of questionable sources ($\alpha \sim 1/2$), while Gab strongly amplifies them ($\alpha \sim 4$).

Therefore, we conclude that the main drivers of information spreading are related to specific peculiarities of each platform and depends upon the group dynamics of individuals engaged with the topic.

6.3 Conclusions

In this work we perform a comparative analysis of users' activity on five different social media platforms during the COVID-19 health emergency. Such a timeframe is a good benchmark for studying content consumption dynamics around critical events in a times when the accuracy of information is threatened. We assess user engagement and interest about the COVID-19 topic and characterize the evolution of the discourse over time.

Furthermore, we model the spread of information using epidemic models and provide basic growth parameters for each social media platform. We then analyze the diffusion of questionable information for all channels, finding that Gab is the environment more susceptible to misinformation dissemination. However, information deriving from sources marked either as reliable or questionable do not present significant differences in their its spreading patterns. Our analysis suggests that information spreading is driven by the interaction paradigm imposed by the specific social media or/and by the specific interaction patterns of groups of users engaged with the topic. We conclude the paper by computing rumor's amplification parameters for social media platforms.

We believe that the understanding of social dynamics between content consumption and social media platforms is an important research subject, since it may help to design more efficient epidemic models accounting for social behavior and to design more effective and tailored communication strategies in time of crisis.

6.4 Methods

6.4.1 Data Collection

Table 6.3 reports the data breakdown of the five social media platforms. Different data collection processes have been performed depending on the platform. In all cases we guided the data collection by a set of selected keywords based on Google Trends' COVID-19 related queries such as: coronavirus, pandemic, coronaoutbreak, china, wuhan, nCoV, IamNotAVirus, coronavirus_update, coronavirus_transmission, coronavirusnews, coronavirusoutbreak.

The Reddit dataset was downloaded from the Pushift.io archive, exploiting the related API. In order to filter contents linked to COVID-19, we used our set of keywords.

In Gab, although no official guides are available, there is an API service that given a certain keyword, returns a list of users, hashtags and groups related to it. We queried all the keywords we selected based on Google Trends and we downloaded all hashtags linked to them. We then manually browsed the results and selected a set of hashtags based on their meaning. For each hashtag in our list, we downloaded all the posts and comments linked to it.

For YouTube, we collected videos by using the YouTube Data API by searching for videos that matched our keywords. Then an in depth search was done by crawling the network of videos by searching for more related videos as established by the YouTube algorithm. From the gathered set, we filtered the videos that matched coronavirus, nCov, corona virus, corona-virus, corvid, covid or SARS-CoV in the title or description. We then collected all the comments received by those videos.

For Twitter, we collect tweets related to the topic coronavirus by using both the search and stream endpoint of the Twitter API. The data derived from the stream API represent only 1% of the total volume of tweets, further filtered by the selected keywords. The data derived from the search API represent a random sample of the tweets containing the selected keywords up to a maximum rate limit of 18000 tweets every 10 minutes.

Since no official API are available for Instagram data, we built our own process to collect public contents related to our keywords. We manually took notes of posts, comments and populated the Instagram Dataset.

	Posts	Comments	Users	Period
Gab	6,252	4,364	2,629	01/01-14/02
Reddit	10,084	300,751	89,456	01/01-14/02
YouTube	111,709	7,051,595	3,199,525	01/01-14/02
Instagram	26,576	109,011	52,339	01/01-14/02
Twitter	1,187,482	-	390,866	27/01-14/02
Total	1,342,103	7,465,721	3,734,815	

Table 6.3: Data breakdown of the number of posts, comments and users for all platforms.

6.4.2 Matching Ability

We consider all the posts in our dataset that contain at least one URL linking to a website outside the related social media platfrom (e.g., tweets pointing outside Twitter). We separate URLs in two main categories obtained using the classification provided by MediaBias/FactCheck (MBFC). MBFC provides a classification determined by ranking bias in four different categories, one of them being Factual/Sourcing. In that category, each news outlet is associated to a label that refers to its reliability as expressed in three labels, namely Conspiracy-Pseudoscience, Pro-Science or Questionable. Noticeably, also the Questionable set include a wide range of political bias, from Extreme Left to Extreme Right.

Using such a classification, we assign to each of these outlets a binary label that partially stems from the labelling provided by MBFC. We divide the news outlets into Questionable and Reliable. All the outlets already classified as Questionable or belonging to the category Conspiracy-Pseudoscience are labelled as Question-

	Gab	Reddit	YouTube	Instagram	Twitter
Posts containing a URL	3778	10084	351786	1328	356448
Matched	0.47	0.55	0.035	0.09	0.27
Questionable	0.38	0.045	0.064	0.05	0.10
Reliable	0.62	0.955	0.936	0.95	0.90

Table 6.4: Number of posts containing a URL, matching ability and classification for each of the five platforms.

able, the rest is labelled as Reliable. Thus, by questionable information source we mean a news outlet systematically showing one or more of the following characteristics: extreme bias, consistent promotion of propaganda/conspiracies, poor or no sourcing to credible information, information not supported by evidence or unverifiable, a complete lack of transparency and/or fake news. By reliable information sources we mean news outlets that do not show any of the aforementioned characteristics. Such outlets can anyway produce contents potentially displaying a bias towards liberal/conservative opinion, but this does not compromise the overall reliability of the source.

Considering all the 2637 news outlets that we retrieve from the list provided by MBFC we end up with 800 outlets classified as Questionable 1837 outlets classified as Reliable. Using such a classification we quantify our overall ability to match and label domains of posts containing URLs, as reported in Table 6.4. The matching ability that is low doesn't refer to the ability of identifying known domain but to the ability of finding the news outlets that belong to the list provided by MBFC. Indeed in all the social networks we find a tendency towards linking to other social media platforms, as shown in Table 6.5.

	Gab	Reddit	YouTube	Instagram	Twitter	Facebook
Gab	0.003	0.002	0.001	0.002	0.138	$\sim$0
Reddit	0.043	0.006	0.009	0.001	$\sim$0	0
YouTube	0	$\sim$0	0.292	$\sim$0	0.088	0.081
Instagram	0	0	0.003	0	0.001	0.001
Twitter	0.059	0.001	0.257	0.003	$\sim$0	$\sim$0

Table 6.5: Fraction of URLs pointing to social media. Table should be read as entries in each row link to entries in each column. For example, Gab links to Reddit 0.003.

6.4.3 Text Analysis

To provide an overview of the debate concerning the virus outbreak on the various platforms, we extract and analyze all topics related to COVID-19 by applying Natural Language Processing techniques to the written content of each social media platform. We first build word embedding for the text corpus of each platform, then, to assess the topics around which the perception of the COVID-19 debate is concentrated, we cluster words by running the Partitioning Around Medoids (PAM) algorithm on their vector representations.

Word embeddings, i.e., distributed representations of words learned by neural networks, represent words as vectors in $\mathbf{R}^n$ bringing similar words closer to each other. They perform significantly better than the well-known Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA) for preserving linear regularities among words and computational efficiency on large data sets [175]. In this paper we use the Skip-gram model [51] to construct word embedding of each social media corpus. More formally, given a content represented by the sequence of words $w_1, w_2, \ldots, w_T$, we use stochastic gradient descent with gradient computed through backpropagation rule [52] for maximizing the average log probability

$$\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{j=-k}^{k}\log p(w_{t+j}|w_t)\right] \tag{6.1}$$

where k is the size of the training window. Therefore, during training the vector representations of closely related words are pushed to be close to each other.

In the Skip-gram model, every word w is associated with its input and output vectors, u_w and v_w, respectively. The probability of correctly predicting the word w_i given the word w_j is defined as

$$p(w_i|w_j) = \frac{\exp\left(u_{w_i}^T v_{w_j}\right)}{\sum_{l=1}^{V}\exp\left(u_l^T v_{w_j}\right)} \tag{6.2}$$

where V is the number of words in the corpus vocabulary. Two major parameters affect the training quality: the dimensionality of word vectors, and the size of the surrounding words window. We choose 200 as vector dimension – that is typical value for training large dataset – and 6 words for the window.

Before applying the tool, we reduced the contents to those written in English as detected with cld3. Then we cleaned the corpora by removing HTML code, URLs and email addresses, user mentions, hashtags, stop-words, and all the special characters including digits. Finally, we dropped words composed by less than three characters, words occurring less than five times in all the corpus, and contents with less than three words.

To analyze the topics related to COVID-19, we cluster words by PAM and using as proximity metric the cosine distance matrix of words in their vector representations. In order to select the number of clusters, k, we calculate the average silhouette width for each value of k. Moreover, for evaluating the cluster stability, we calculate the average pairwise Jaccard similarity between clusters based on 90% sub-samples of the data. Lastly, we produce word clouds to identify the topic of each cluster. To provide a view about the debate around the virus outbreak, we define the distribution over topics Θ_c for a given content c as the distribution of its words among the word clusters. Thus, to quantify the relevance of each topic within a corpus, we restrict to contents c with $\max \Theta_c > 0.5$ and consider them uniquely identified as a single topic each. Table 6.6 shows the results of the text cleaning and topic analysis for all the data.

	Cleaned contents	**Vocabulary size**	**Topics**	**Contents with** $\max\Theta > 0.5$
Instagram	21,189 posts	15,324	17	4,467
Twitter	638,214 posts	22,587	21	369,131
Gab	5,853 posts	3,024	19	2,986
Reddit	10,084 posts	1,968	34	6,686
YouTube	815,563 comments	35,381	30	679,261

Table 6.6: Results of text cleaning and analysis for all the corpora.

6.4.4 Epidemiological Models

Several mathematical models can be used to analyse potential mechanisms that underline epidemiological data. Generally, we can distinguish among phenomenological models that emphasize the reproducibility of empirical data without insights in the mechanisms of growth, and more insightful mechanistic models that try to incorporate such mechanisms [172].

To fit our cumulative curves, we first use the adjusted exponential model of [53] since it naturally provides an estimate of the basic reproduction number R_0. This phenomenological model (from now on indicated as EXP) has been successfully employed in data-scarce settings and shown to be on-par with more traditional compartmental models for multiple emerging diseases like Zika, Ebola, and Middle East Respiratory Syndrome [53].

The model is defined by the following single equation:

$$I = \left[\frac{R_0}{(1+d)^t}\right]^t \tag{6.3}$$

Here, I is incidence, t is the number of days, R_0 is the basic reproduction number and d is a damping factor accounting for the reduction in transmissibility over time. In our case, we interpret I as the number C_{auth} of authors that have published a post on the subject.

As a mechanistic model, we employ the classical SIR model [54]. In such a model, a susceptible population can be infected with a rate β by coming into contact with infected individuals; however, infected individuals can recover with a rate γ. The model is described by a set of differential equations:

$$\begin{aligned} \partial_t S &= -\beta S \cdot I/N \\ \partial_t I &= \beta S \cdot I/N - \gamma I \\ \partial_t R &= \gamma I \end{aligned} \tag{6.4}$$

where S is the number of susceptible, I is the number of infected and R is the number of recovered. In our case, we interpret the number $I + R$ as the number C_{auth} of authors that have published a post on the subject.

In the SIR model, the basic reproduction number $R_0 = \beta/\gamma$ corresponds to the ration among the rate of infection by contact β and the rate of recovery γ. Notice that for the SIR model, vaccination strategies correspond to bringing the system in a situation where $S < N/R_0$; in such a way, both the number of infected will decrease.

To estimate the basic reproduction numbers R_0^{EXP} and R_0^{SIR} for the EXP and the SIR model, we use least square estimates of the models' parameters[55]. The range of parameters is estimated via bootstrapping [172], [176].

6.4.5 Linear Regression Coefficients

Table 6.7 reports the regression coefficient ρ, the intercept and the R^2 values for the linear fit of Figure 6.3. High values of R^2 confirm the linear relationship between reliable and questionable sources in information diffusion.

Dataset	Type	Intercept	Coefficient (ρ)	R^2
Gab	Posts	-22.321	0.695	0.996
Reddit	Posts	-4.111	0.047	0.997
Youtube	Posts	4.529	0.073	0.998
Twitter	Posts	-151.44	0.110	0.998
Gab	Reactions	74.577	2.721	0.981
Reddit	Reactions	-70.677	0.026	0.990
Youtube	Reactions	-8854.33	0.025	0.986
Twitter	Reactions	-2136.978	0.107	0.987

Table 6.7: Coefficients and R^2 of the linear regressions displayed in Figure 3.

CHAPTER 7

CONCLUSION AND FUTURE WORKS

In this chapter, we summarize the results of the works presented, drawn conclusions, and sketch the line of the research for future works.

7.1 Summary of Key Findings

The research presented in this thesis mainly focused on two aspects of the social media environment: information consumption dynamics and echo chambers. We first showed the marginal role of fake news during the 2019 European elections, then we analyzed and quantified the difference in polarization between the 2016 and 2020 U.S. presidential elections. We also compared the share of misinformation and fake news circulating, showing the decrease in the presence of fake news and automated accounts. We also highlighted the presence of two clusters of users with opposite political leaning that increased the opinion distance over time. Thus, we looked at the possible environmental factors that may foster the polarization of users and the rise of echo chambers by comparing the debate around several topics on different platforms. We found that social media feed algorithms may foster the echo chamber effects and that information consumption is influenced by the structure of echo chambers. Finally, we observed how different social media platforms reacted to the COVID-19 outbreak. We found a dramatic increase in the amount of news shared on all platforms that leads to an overabundance of both reliable and questionable information. This uncontrolled proliferation of COVID-19 contents has been so impressive to the point that the term “infodemic” has been coined to describe it. Moreover, we compared the proliferation of questionable and reliable news among different platforms finding that their level of diffusion and consumption depends on the environment in which they spread, but has comparable dynamics for both types of news. To conclude, the level of diffusion of a piece of information depends on the presence of an audience, or better an echo chamber, prone to endorse that type of content. However, several factors such as the platform feature, the characteristic of the user base or the feed algorithm, may influence the existence of such users and thus the level of diffusion of the information. Nevertheless, the spreading dynamic seems to be independent from some characteristic of the content, such as reliability.

7.2 Future Works

Although a considerable amount of research about echo chambers and information spreading has been published, there are still open questions. One of the most important is the process dominating the rise of polarization and echo chambers.

We now have several techniques to reveal the presence of polarization and echo chambers, but we still lack tools to describe the rise and evolution of echo chambers. However, this line of research is likely to benefit from the vast amount of historical data that some platforms recently made available for academic purposes. A better understanding of the echo chambers evolution may aid the design of actions and feed algorithms that reduce polarization and segregation among users. Another important question is the quantification of the segregation level among online communities. Even if some studies tried to model polarization, we still miss a metric to measure the polarization level for different topics and across several platforms, allowing us to compare and study the evolution of polarization over time. A crucial element of online social media studies is the representativeness and the biases that possibly affect the data gathered from online platforms. Indeed, online data may not reflect the offline world in an unbiased manner or not adequately represent all the aspects of an online environment, and the effects on research results of such biases still need to be precisely addressed. Finally, we still lack of knowledge about the relationship between online debate and offline actions. Frequently, social media platforms have been identified as the trigger of actions such as assault, shootings and radicalization, but we still do not know to which extent the online echo system can influence offline behaviours.

REFERENCES

[1] W. Quattrociocchi, "Part 2-social and political challenges: 2.1 western democracy in crisis?" In *Global Risk Report World Economic Forum*, 2017.

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

[3] A. L. Schmidt, F. Zollo, M. Del Vicario, A. Bessi, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, "Anatomy of news consumption on facebook," *Proceedings of the National Academy of Sciences*, vol. 114, no. 12, pp. 3035–3039, 2017.

[4] M. Cinelli, E. Brugnoli, A. L. Schmidt, F. Zollo, W. Quattrociocchi, and A. Scala, "Selective exposure shapes the facebook news diet," *PloS one*, vol. 15, no. 3, e0229129, 2020.

[5] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.

[6] K. H. Jamieson and J. N. Cappella, *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press, 2008.

[7] R. K. Garrett, "Echo chambers online?: Politically motivated selective exposure among Internet news users," *Journal of Computer-Mediated Communication*, vol. 14, no. 2, pp. 265–285, 2009.

[8] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, "Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18, Lyon, France: International World Wide Web Conferences Steering Committee, 2018, pp. 913–922, ISBN: 978-1-4503-5639-8.

[9] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, "The Effect of Collective Attention on Controversial Debates on Social Media," in *WebSci '17: 9th International ACM Web Science Conference*, 2017, pp. 43–52.

[10] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini, “Quantifying echo chamber effects in information spreading over political communication networks,” *EPJ Data Science*, vol. 8, no. 1, p. 35, Dec. 2019.

[11] E. Dubois and G. Blank, “The echo chamber is overstated: The moderating effect of political interest and diverse media,” *Information, Communication & Society*, vol. 21, no. 5, pp. 729–745, 2018.

[12] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, “Tweeting from left to right: Is online political communication more than an echo chamber?” *Psychological science*, vol. 26, no. 10, pp. 1531–1542, 2015.

[13] A. Bruns, *Echo chamber? what echo chamber? reviewing the evidence*, 2017.

[14] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[15] C. Shao, G. L. Ciampaglia, O. Varol, A. Flammini, and F. Menczer, “The spread of fake news by social bots,” *arXiv preprint arXiv:1707.07592*, vol. 96, pp. 96–104, 2017.

[16] D. Ruths, “The misinformation machine,” *Science*, vol. 363, no. 6425, pp. 348–348, 2019.

[17] A. Bessi and E. Ferrara, “Social bots distort the 2016 us presidential election online discussion,” *First Monday*, vol. 21, no. 11-7, 2016.

[18] A. Bovet and H. A. Makse, “Influence of fake news in twitter during the 2016 us presidential election,” *Nature communications*, vol. 10, no. 1, p. 7, 2019.

[19] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, “Fake news on twitter during the 2016 us presidential election,” *Science*, vol. 363, no. 6425, pp. 374–378, 2019.

[20] *Twitter stops all political advertising*, `https://twitter.com/jack/status/1189634360472829952`, Accessed: 2020-04-04.

[21] *Is russia trying to sway the european elections?* `https://www.bbc.com/news/blogs-trending-48296557`, Accessed: 2020-04-04.

[22] *Fakewatch*, https://secure.avaaz.org/campaign/en/disinfo_network_report/, Accessed: 2020-04-04.

[23] *Eu leaders to warn of cyber, fake news threat to may elections*, https://www.reuters.com/article/us-eu-disinformation/eu-leaders-to-warn-of-cyber-fake-news-threat-to-may-elections-idUSKCN1R11QM, Accessed: 2020-04-04.

[24] J. Zarocostas, "How to fight an infodemic," *The Lancet*, vol. 395, no. 10225, p. 676, 2020.

[25] *Why social media makes us more polarized and how to fix it*, https://www.scientific-american.com/article/why-social-media-makes-us-more-polarized-and-how-to-fix-it/, Accessed: 2021-09-15.

[26] *A mass murder of, and for, the internet*, https://www.nytimes.com/2019/03/15/technology/facebook-youtube-christchurch-shooting.html, Accessed: 2021-09-15.

[27] *On gab, an extremist-friendly site, pittsburgh shooting suspect aired his hatred in full*, https://www.nytimes.com/2018/10/28/us/gab-robert-bowers-pittsburgh-synagogue-shootings.html, Accessed: 2021-09-15.

[28] *Far-right extremists still downloading christchurch massacre footage*, https://www.theage.com.au/national/far-right-extremists-still-downloading-christchurch-massacre-footage-20210818-p58jwq.html, Accessed: 2021-09-15.

[29] *The storming of capitol hill was organized on social media.* https://www.nytimes.com/2021/01/06/us/politics/protesters-storm-capitol-hill-building.html, Accessed: 2021-09-15.

[30] *Capitol hill riot lays bare what's wrong with social media*, https://www.politico.eu/article/us-capitol-hill-riots-lay-bare-whats-wrong-social-media-donald-trump-facebook-twitter/, Accessed: 2021-09-15.

[31] *The 65 days that led to chaos at the capitol*, https://www.bbc.com/news/world-us-canada-55592332, Accessed: 2021-09-15.

[32] *Permanent suspension of @realdonaldtrump*, https://about.fb.com/news/2021/01/responding-to-the-violence-in-washington-dc/, Accessed: 2021-09-15.

[33] *Permanent suspension of @realdonaldtrump*, https://blog.twitter.com/en_us/topics/company/2020/suspension, Accessed: 2021-09-15.

[34] *Trump sues twitter, google and facebook alleging 'censorship'*, https://www.bbc.com/news/world-us-canada-57754435, Accessed: 2021-09-15.

[35] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, *et al.*, "Computational social science.," *Science (New York, NY)*, vol. 323, no. 5915, pp. 721–723, 2009.

[36] M. Baker, "Over half of psychology studies fail reproducibility test," *Nature News*, vol. 27, 2015.

[37] B. Owens, "Replication failures in psychology not due to differences in study populations," *Nature*, vol. 19, 2018.

[38] B. Gonçalves, N. Perra, and A. Vespignani, "Modeling users' activity on twitter networks: Validation of dunbar's number," *PloS one*, vol. 6, no. 8, e22656, 2011.

[39] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–36, 2017.

[40] A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Science vs conspiracy: Collective narratives in the age of misinformation," *PloS one*, vol. 10, no. 2, e0118093, 2015.

[41] M. Del Vicario, S. Gaito, W. Quattrociocchi, M. Zignani, and F. Zollo, "News consumption during the italian referendum: A cross-platform analysis on facebook and twitter," in *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, 2017, pp. 648–657.

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

[43] P. Barberá, *Social media, echo chambers, and political polarization*, 2020.

[44] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, “Quantifying Controversy in Social Media,” in *WSDM ’16: 9th ACM International Conference on Web Search and Data Mining*, 2016, pp. 33–42.

[45] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, P10008, 2008.

[46] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, “On modularity clustering,” *IEEE transactions on knowledge and data engineering*, vol. 20, no. 2, pp. 172–188, 2007.

[47] P. Barberá, “Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data,” *Political analysis*, vol. 23, no. 1, pp. 76–91, 2015.

[48] J.-P. Benzécri *et al.*, *L’analyse des données*. Dunod Paris, 1973, vol. 2.

[49] O. Nenadic and M. Greenacre, “Correspondence analysis in r, with two-and three-dimensional graphics: The ca package,” *Journal of statistical software*, vol. 20, no. 3, 2007.

[50] P. Holme, “Network reachability of real-world contact sequences,” *Phys. Rev. E*, vol. 71, p. 046 119, 4 Apr. 2005.

[51] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS’13, Lake Tahoe, Nevada: Curran Associates Inc., 2013, pp. 3111–3119.

[52] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[53] D. N. Fisman, T. S. Hauck, A. R. Tuite, and A. L. Greer, “An idea for short term outbreak projection: Nearcasting using the basic reproduction number,” *PloS one*, vol. 8, no. 12, 2013.

[54] N. T. Bailey *et al.*, *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

[55] J. Ma, "Estimating epidemic exponential growth rate and basic reproduction number," *Infectious Disease Modelling*, 2020.

[56] F. Zollo and W. Quattrociocchi, "Misinformation spreading on facebook," in *Complex Spreading Phenomena in Social Systems*, Springer, 2018, pp. 177–196.

[57] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Echo chambers: Emotional contagion and group polarization on facebook," *Scientific reports*, vol. 6, p. 37 825, 2016.

[58] V. Bakir and A. McStay, "Fake news and the economy of emotions: Problems, causes, solutions," *Digital journalism*, vol. 6, no. 2, pp. 154–175, 2018.

[59] D. A. Scheufele and N. M. Krause, "Science audiences, misinformation, and fake news," *Proceedings of the National Academy of Sciences*, vol. 116, no. 16, pp. 7662–7669, 2019.

[60] W. Quattrociocchi, G. Caldarelli, and A. Scala, "Opinion dynamics on interacting networks: Media competition and social influence," *Scientific reports*, vol. 4, p. 4938, 2014.

[61] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, "Fake news on twitter during the 2016 u.s. presidential election," *Science*, vol. 363, no. 6425, pp. 374–378, 2019. eprint: `https://science.sciencemag.org/content/363/6425/374.full.pdf`.

[62] R. Wald, T. M. Khoshgoftaar, A. Napolitano, and C. Sumner, "Predicting susceptibility to social bots on twitter," in *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)*, IEEE, 2013, pp. 6–13.

[63] N. Mele, D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph, W. Hobbs, and C. Mattsson, "Combating fake news: An agenda for research and action," *Di https://www.hks.harvard.edu/publications/combating-fake-news-agenda-research-and-action (Retrieved October 17, 2018)*, 2017.

[64] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proceedings of the 26th International Conference on World Wide Web Companion (WWW'17 Companion)*, 2017, pp. 963–972.

[65] A. Fourney, M. Z. Racz, G. Ranade, M. Mobius, and E. Horvitz, "Geographic and temporal trends in fake news consumption during the 2016 us presidential election," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ACM, 2017, pp. 2071–2074.

[66] S. Cresci, F. Lillo, D. Regoli, S. Tardelli, and M. Tesconi, "Cashtag piggybacking: Uncovering spam and bot activity in stock microblogs on twitter," *ACM Transactions on the Web (TWEB)*, vol. 13, no. 2, p. 11, 2019.

[67] M. D. Vicario, F. Zollo, G. Caldarelli, A. Scala, and W. Quattrociocchi, "The anatomy of brexit debate on facebook," *CoRR*, vol. abs/1610.06809, 2016. arXiv: `1610.06809`.

[68] M. D. Vicario, W. Quattrociocchi, A. Scala, and F. Zollo, "Polarization and fake news: Early warning of potential misinformation targets," *ACM Transactions on the Web (TWEB)*, vol. 13, no. 2, pp. 1–22, 2019.

[69] W. Quattrociocchi, "Inside the echo chamber," *Scientific American*, vol. 316, no. 4, pp. 60–63, 2017.

[70] *Tackling online disinformation*, `https://secure.avaaz.org/campaign/en/disinfo_network_report/`, Accessed: 2020-04-04.

[71] F. Pierri, A. Artoni, and S. Ceri, "Investigating italian disinformation spreading on twitter in the context of 2019 european elections," *PloS one*, vol. 15, no. 1, e0227821, 2020.

[72] M. Stella, E. Ferrara, and M. De Domenico, "Bots increase exposure to negative and inflammatory content in online social systems," *Proceedings of the National Academy of Sciences*, vol. 115, no. 49, pp. 12 435–12 440, 2018. eprint: `https://www.pnas.org/content/115/49/12435.full.pdf`.

[73] *Measuring the reach of "fake news" and online disinformation in europe*, `https://reutersinstitute.politics.ox.ac.uk/our-research/measuring-`

reach-fake-news-and-online-disinformation-europe, Accessed: 2020-04-04.

[74] *European journalism observatory*, https://en.ejo.ch/, Accessed: 2020-04-04.

[75] *Snopes*, https://www.snopes.com/, Accessed: 2020-04-04.

[76] *Politifact*, https://www.politifact.com/, Accessed: 2020-04-04.

[77] *European media monitor*, https://emm.newsbrief.eu/overview.html, Accessed: 2020-04-04.

[78] *Members of the european parliament*, http://www.europarl.europa.eu/meps/en/home, Accessed: 2020-04-04.

[79] D. W. Brady and H. C. Han, "Polarization then and now: A historical perspective," *Red and blue nation*, vol. 1, pp. 119–151, 2006.

[80] C. Hare and K. T. Poole, "The polarization of contemporary american politics," *Polity*, vol. 46, no. 3, pp. 411–429, 2014.

[81] R. Axelrod, J. J. Daymude, and S. Forrest, "Preventing extreme polarization of political attitudes," *arXiv preprint arXiv:2103.06492*, 2021.

[82] D. Guilbeault and D. Centola, "Topological measures for identifying and predicting the spread of complex contagions," *Nature communications*, vol. 12, no. 1, pp. 1–9, 2021.

[83] P. Metaxas, E. Mustafaraj, K. Wong, L. Zeng, M. O'Keefe, and S. Finn, "What do retweets indicate? results from user survey and meta-review of research," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 9, 2015.

[84] N. McCarty, *Polarization: What everyone needs to know®*. Oxford University Press, 2019.

[85] W. A. Galston and P. S. Nivola, "Delineating the problem," *Red and blue nation*, vol. 1, pp. 1–47, 2006.

[86] A. I. Abramowitz and M. P. Fiorina, “Polarized or sorted? just what’s wrong with our politics, anyway,” *The American Interest*, 2013.

[87] M. P. Fiorina and S. J. Abrams, “Political polarization in the american public,” *Annu. Rev. Polit. Sci.*, vol. 11, pp. 563–588, 2008.

[88] G. C. Layman, T. M. Carsey, and J. M. Horowitz, “Party polarization in american politics: Characteristics, causes, and consequences,” *Annu. Rev. Polit. Sci.*, vol. 9, pp. 83–110, 2006.

[89] L. Mason, “‘‘i disrespectfully agree”: The differential effects of partisan sorting on social and issue polarization,” *American Journal of Political Science*, vol. 59, no. 1, pp. 128–145, 2015.

[90] R. Effing, J. Van Hillegersberg, and T. Huibers, “Social media and political participation: Are facebook, twitter and youtube democratizing our political systems?” In *International conference on electronic participation*, Springer, 2011, pp. 25–35.

[91] M. Broersma and T. Graham, “Social media as beat: Tweets as a news source during the 2010 british and dutch elections,” *journalism practice*, vol. 6, no. 3, pp. 403–419, 2012.

[92] P. T. Metaxas and E. Mustafaraj, “Social media and the elections,” *Science*, vol. 338, no. 6106, pp. 472–473, 2012.

[93] A. Ceron, L. Curini, and S. M. Iacus, *Politics and big data: Nowcasting and forecasting elections with social media*. Routledge, 2016.

[94] A. Bovet, F. Morone, and H. A. Makse, “Validation of twitter opinion trends with national polling aggregates: Hillary clinton vs donald trump,” *Scientific reports*, vol. 8, no. 1, pp. 1–16, 2018.

[95] F. B. Soares, R. Recuero, and G. Zago, “Influencers in polarized political networks on twitter,” in *Proceedings of the 9th international conference on social media and society*, 2018, pp. 168–177.

[96] P. Grover, A. K. Kar, Y. K. Dwivedi, and M. Janssen, “Polarization and acculturation in us election 2016 outcomes–can twitter analytics predict changes in voting preferences,” *Technological Forecasting and Social Change*, vol. 145, pp. 438–460, 2019.

[97] S. Lee and M. Xenos, “Social distraction? social media use and political knowledge in two us presidential elections,” *Computers in human behavior*, vol. 90, pp. 18–25, 2019.

[98] Z. Acharoui, A. Alaoui, B. Ettaki, J. Zerouaoui, and M. Dakkon, “Identifying political influencers on youtube during the 2016 moroccan general election,” *Procedia Computer Science*, vol. 170, pp. 1102–1109, 2020.

[99] G. Suau-Gomila, C. Pont-Sorribes, and R. Pedraza-Jiménez, “Politicians or influencers? twitter profiles of pablo iglesias and albert rivera in the spanish general elections of 20-d and 26-j,” *Communication & Society*, pp. 209–225, 2020.

[100] C. Shao, P.-M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczer, and G. L. Ciampaglia, “Anatomy of an online misinformation network,” *PloS one*, vol. 13, no. 4, e0196087, 2018.

[101] C. Machado, B. Kira, V. Narayanan, B. Kollanyi, and P. Howard, “A study of misinformation in whatsapp groups with a focus on the brazilian presidential elections.,” in *Companion proceedings of the 2019 World Wide Web conference*, 2019, pp. 1013–1019.

[102] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, “Political polarization on twitter,” in *Fifth international AAAI conference on weblogs and social media*, 2011.

[103] M. Prior, “Media and political polarization,” *Annual Review of Political Science*, vol. 16, pp. 101–127, 2013.

[104] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi, “Collective attention in the age of (mis) information,” *Computers in Human Behavior*, vol. 51, pp. 1198–1204, 2015.

[105] A. Bessi, F. Petroni, M. Del Vicario, F. Zollo, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi, “Homophily and polarization in the age of misinformation,” *The European Physical Journal Special Topics*, vol. 225, no. 10, pp. 2047–2059, 2016.

[106] C. Vaccari, A. Valeriani, P. Barberá, J. T. Jost, J. Nagler, and J. A. Tucker, “Of echo chambers and contrarian clubs: Exposure to political disagreement among german and italian users of twitter,” *Social media+ society*, vol. 2, no. 3, p. 2 056 305 116 664 221, 2016.

[107] A. Bessi, F. Zollo, M. Del Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, and W. Quattrociocchi, "Users polarization on facebook and youtube," *PloS one*, vol. 11, no. 8, e0159641, 2016.

[108] Y. Lelkes, G. Sood, and S. Iyengar, "The hostile audience: The effect of access to broadband internet on partisan affect," *American Journal of Political Science*, vol. 61, no. 1, pp. 5–20, 2017.

[109] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences*, vol. 115, no. 37, pp. 9216–9221, 2018.

[110] M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, "The echo chamber effect on social media," *Proceedings of the National Academy of Sciences*, vol. 118, no. 9, 2021.

[111] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *ACM Transactions on Information Systems*, vol. 28, no. 4, p. 20, 2010.

[112] F. Morone and H. A. Makse, "Influence maximization in complex networks through optimal percolation," *Nature*, vol. 524, no. 7563, pp. 65–68, 2015.

[113] F. Morone, B. Min, L. Bo, R. Mari, and H. A. Makse, "Collective influence algorithm to find influencers via optimal percolation in massively large social media," *Scientific reports*, vol. 6, no. 1, pp. 1–11, 2016.

[114] X. Teng, S. Pei, F. Morone, and H. A. Makse, "Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks," *Scientific reports*, vol. 6, no. 1, pp. 1–11, 2016.

[115] J. A. Hartigan and P. M. Hartigan, "The dip test of unimodality," *The annals of Statistics*, pp. 70–84, 1985.

[116] L. Bozarth, A. Saraf, and C. Budak, "Higher ground? how groundtruth labeling impacts our understanding of fake news about the 2016 us presidential nominees," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, 2020, pp. 48–59.

[117] T. J. Main, *The rise of the alt-right*. Brookings Institution Press, 2018.

[118] P. Stefanov, K. Darwish, A. Atanasov, and P. Nakov, “Predicting the topical stance and political leaning of media using tweets,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 527–537.

[119] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala, “The covid-19 social media infodemic,” *Scientific Reports*, vol. 10, 2020.

[120] L. Bode, “Political news in the news feed: Learning politics from social media,” *Mass communication and society*, vol. 19, no. 1, pp. 24–48, 2016.

[121] N. Newman, R. Fletcher, A. Kalogeropoulos, and R. Nielsen, *Reuters institute digital news report 2019*. Reuters Institute for the Study of Journalism, 2019, vol. 2019.

[122] S. Flaxman, S. Goel, and J. M. Rao, “Filter bubbles, echo chambers, and online news consumption,” *Public opinion quarterly*, vol. 80, no. S1, pp. 298–320, 2016.

[123] T. Sharot and C. R. Sunstein, “How people decide what they want to know,” *Nature Human Behaviour*, vol. 4, pp. 1–6, 2020.

[124] A. Baronchelli, “The emergence of consensus: A primer,” *Royal Society open science*, vol. 5, no. 2, p. 172 189, 2018.

[125] J. T. Klapper, *The effects of mass communication.* 1960.

[126] R. S. Nickerson, “Confirmation bias: A ubiquitous phenomenon in many guises,” *Review of general psychology*, vol. 2, no. 2, pp. 175–220, 1998.

[127] E. Gilbert, T. Bergstrom, and K. Karahalios, “Blogs are echo chambers: Blogs are echo chambers,” in *42nd Hawaii International Conference on System Sciences*, 2009, pp. 1–10.

[128] A. Edwards, “(how) do participants in online discussion forums create ‘echo chambers’?: The inclusion and exclusion of dissenting voices in an online forum about climate change,” *Journal of Argumentation in Context*, vol. 2, no. 1, pp. 127–150, 2013.

[129] M. Grömping, "'Echo Chambers' Partisan Facebook Groups during the 2014 Thai Election," *Asia Pacific Media Educator*, vol. 24, no. 1, pp. 39–59, 2014.

[130] W. Quattrociocchi, A. Scala, and C. R. Sunstein, *Echo chambers on Facebook*, 2016.

[131] I. Himelboim, S. McCreery, and M. Smith, "Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on twitter," *Journal of computer-mediated communication*, vol. 18, no. 2, pp. 154–174, 2013.

[132] D. Nikolov, D. F. Oliveira, A. Flammini, and F. Menczer, "Measuring online social bubbles," *PeerJ Computer Science*, vol. 1, e38, 2015.

[133] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, "Modeling echo chambers and polarization dynamics in social networks," *Physical Review Letters*, vol. 124, no. 4, p. 048 301, 2020.

[134] A. Gollwitzer, C. Martel, W. J. Brady, P. Pärnamets, I. G. Freedman, E. D. Knowles, and J. J. Van Bavel, "Partisan differences in physical distancing are linked to health outcomes during the covid-19 pandemic," *Nature Human Behaviour*, vol. 4, 2020.

[135] Y. Golovchenko, C. Buntain, G. Eady, M. A. Brown, and J. A. Tucker, "Cross-platform state propaganda: Russian trolls on twitter and youtube during the 2016 us presidential election," *The International Journal of Press/Politics*, vol. 25, p. 1 940 161 220 912 682, 2020.

[136] S. Zannettou, B. Bradlyn, E. De Cristofaro, H. Kwak, M. Sirivianos, G. Stringini, and J. Blackburn, "What is gab: A bastion of free speech or an alt-right echo chamber," in *Companion Proceedings of the The Web Conference 2018*, International World Wide Web Conferences Steering Committee, 2018, pp. 1007–1014.

[137] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, "Quantifying controversy on social media," *TSC: ACM Transactions on Social Computing*, vol. 1, no. 1, p. 3, 2018.

[138] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[139] G. Kossinets and D. J. Watts, "Origins of homophily in an evolving social network," *American journal of sociology*, vol. 115, no. 2, pp. 405–450, 2009.

[140] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, "Reducing Controversy by Connecting Opposing Views," in *WSDM '17: 10th ACM International Conference on Web Search and Data Mining*, 2017, pp. 81–90.

[141] R. M. Anderson and R. M. May, *Infectious diseases in humans*. Oxford: Oxford University Press, 1992.

[142] M. Jalili and M. Perc, "Information cascades in complex networks," *Journal of Complex Networks*, vol. 5, no. 5, pp. 665–693, 2017.

[143] L. Zhao, H. Cui, X. Qiu, X. Wang, and J. Wang, "Sir rumor spreading model in the new media age," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 4, pp. 995–1003, 2013.

[144] C. Granell, S. Gómez, and A. Arenas, "Dynamical interplay between awareness and epidemic spreading in multiplex networks," *Phys. Rev. Lett.*, vol. 111, p. 128 701, 12 Sep. 2013.

[145] A. L. Schmidt, F. Zollo, A. Scala, C. Betsch, and W. Quattrociocchi, "Polarization of the vaccination debate on facebook," *Vaccine*, vol. 36, no. 25, pp. 3606–3612, 2018.

[146] H. Lu, J. Caverlee, and W. Niu, "Biaswatch: A lightweight system for discovering and tracking topic-sensitive opinion bias in social media," in *CIKM*, ACM, 2015, pp. 213–222.

[147] W. H. Organization, *Naming the coronavirus disease (covid-19) and the virus that causes it*, `https:www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it`, 2020 (accessed April 9, 2020).

[148] *WHO Situation Report 13*, `https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010_6`, Accessed: 2010-09-30.

[149] W. H. Organization, *Director-general's remarks at the media briefing on 2019 novel coronavirus on 8 february 2020*, `https://www.who.int/dg/speeches/detail/`

director-general-s-remarks-at-the-media-briefing-on-2019-novel-coronavirus---8-february-2020, 2020 (accessed April 9, 2020).

[150] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: Can we trust what we rt?" In *Proceedings of the first workshop on social media analytics*, 2010, pp. 71–79.

[151] K. Starbird, J. Maddock, M. Orand, P. Achterman, and R. M. Mason, "Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing," *IConference 2014 Proceedings*, 2014.

[152] L. Kim, S. M. Fast, and N. Markuzon, "Incorporating media data into a model of infectious disease transmission," *PloS one*, vol. 14, no. 2, 2019.

[153] T. John and C. Ben Wedeman, *Italy prohibits travel and cancels all public events in its northern region to contain coronavirus*, https://edition.cnn.com/2020/03/08/europe/italy-coronavirus-lockdown-europe-intl/index.html, 2020 (accessed April 9, 2020).

[154] J. Shaman, A. Karspeck, W. Yang, J. Tamerius, and M. Lipsitch, "Real-time influenza forecasts during the 2012–2013 season," *Nature communications*, vol. 4, no. 1, pp. 1–10, 2013.

[155] C. Viboud and A. Vespignani, "The future of influenza forecasts," *Proceedings of the National Academy of Sciences*, vol. 116, no. 8, pp. 2802–2804, 2019.

[156] J. Kulshrestha, M. Eslami, J. Messias, M. B. Zafar, S. Ghosh, K. P. Gummadi, and K. Karahalios, "Quantifying search bias: Investigating sources of bias for political searches in social media," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2017, pp. 417–432.

[157] M. Starnini, M. Frasca, and A. Baronchelli, "Emergence of metapopulations and echo chambers in mobile agents," *Scientific reports*, vol. 6, p. 31 834, 2016.

[158] F. Zollo, A. Bessi, M. Del Vicario, A. Scala, G. Caldarelli, L. Shekhtman, S. Havlin, and W. Quattrociocchi, "Debunking in a world of tribes," *PloS one*, vol. 12, no. 7, 2017.

[159] C. Wardle and H. Derakhshan, "Information disorder: Toward an interdisciplinary framework for research and policy making," *Council of Europe report*, vol. 27, 2017.

[160] L. Schild, C. Ling, J. Blackburn, G. Stringhini, Y. Zhang, and S. Zannettou, ""go eat a bat, chang!": An early look on the emergence of sinophobic behavior on web communities in the face of covid-19," *arXiv preprint arXiv:2004.04046*, 2020.

[161] N. Velásquez, R. Leahy, N. J. Restrepo, Y. Lupu, R. Sear, N. Gabriel, O. Jha, and N. Johnson, "Hate multiverse spreads malicious covid-19 content online beyond individual platform control," *arXiv preprint arXiv:2004.00673*, 2020.

[162] E. Ferrara, "What types of covid-19 conspiracies are populated by twitter bots?" *First Monday*, 2020.

[163] F. Alam, S. Shaar, A. Nikolov, H. Mubarak, G. D. S. Martino, A. Abdelali, F. Dalvi, N. Durrani, H. Sajjad, K. Darwish, *et al.*, "Fighting the covid-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society," *arXiv preprint arXiv:2005.00033*, 2020.

[164] G. K. Shahi, A. Dirkson, and T. A. Majchrzak, "An exploratory study of covid-19 misinformation on twitter," *arXiv preprint arXiv:2005.05710*, 2020.

[165] M. B. C. (MBFC), *Media bias/fact check, the most comprehensive meida bias check resource*, `https://mediabiasfactcheck.com/`, 2020 (accessed April 9, 2020).

[166] D. M. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 695–704.

[167] W. H. Organization, *Novel coronavirus (2019-ncov) situation report - 1 21 january 2020*, `https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4`, 2020 (accessed April 9, 2020).

[168] L. Pellis, F. Ball, S. Bansal, K. Eames, T. House, V. Isham, and P. Trapman, "Eight challenges for network epidemic models," *Epidemics*, vol. 10, pp. 58–62, 2015.

[169] Y. Liu, B. Wang, B. Wu, S. Shang, Y. Zhang, and C. Shi, "Characterizing super-spreading in microblog: An epidemic-based information propagation model," *Physica A: Statistical Mechanics and its Applications*, vol. 463, pp. 202–218, 2016.

[170] J. Skaza and B. Blais, "Modeling the infectiousness of twitter hashtags," *Physica A: Statistical Mechanics and its Applications*, vol. 465, pp. 289–296, 2017.

[171] J. T. Davis, N. Perra, Q. Zhang, Y. Moreno, and A. Vespignani, "Phase transitions in information spreading on structured populations," *Nature Physics*, pp. 1–7, 2020.

[172] G. Chowell, "Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts," *Infectious Disease Modelling*, vol. 2, no. 3, pp. 379–398, 2017.

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

[174] M. Del Vicario, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, "Modeling confirmation bias and polarization," *Scientific reports*, vol. 7, p. 40 391, 2017.

[175] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia: Association for Computational Linguistics, Jun. 2013, pp. 746–751.

[176] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.

APPENDIX A

TABLES FOR MEASURING THE EVOLUTION OF POLARIZATION AND NEWS INFLUENCERS BETWEEN TWO U.S. PRESIDENTIAL ELECTIONS ON TWITTER

	fake news		extreme bias (right) news		right news	
	hostnames	N	hostnames	N	hostnames	N
1	thegatewaypundit.com	761 756	breitbart.com	1 854 920	foxnews.com	1 122 732
2	truthfeed.com	554 955	dailycaller.com	759 504	dailymail.co.uk	474 846
3	infowars.com	478 872	americanthinker.com	179 696	washingtonexaminer.com	462 769
4	therealstrategy.com	241 354	wnd.com	141 336	nypost.com	441 648
5	conservativetribune.com	212 273	freebeacon.com	129 077	bizpacreview.com	170 770
6	zerohedge.com	186 706	newsninja2012.com	127 251	nationalreview.com	164 036
7	rickwells.us	78 736	hannity.com	114 221	lifezette.com	139 257
8	departed.co	72 773	newsmax.com	94 882	redstate.com	105 912
9	thepoliticalinsider.com	66 426	endingthefed.com	88 376	allenbwest.com	104 857
10	therightscoop.com	63 852	truepundit.com	84 967	theconservativetreehouse.com	102 515
11	teaparty.org	48 757	westernjournalism.com	77 717	townhall.com	102 408
12	usapoliticsnow.com	46 252	dailywire.com	67 893	investors.com	102 295
13	clashdaily.com	45 970	newsbusters.org	60 147	theblaze.com	99 029
14	thefederalistpapers.org	45 831	ilovemyfreedom.org	54 772	theamericanmirror.com	91 538
15	redflagnews.com	45 423	100percentfedup.com	54 596	ijr.com	71 558
16	thetruthdivision.com	44 486	pjmedia.com	46 542	judicialwatch.org	70 543
17			weaselzippers.us	45 199	thefederalist.com	55 835
18					hotair.com	55 431
19					conservativereview.com	54 307
20					weeklystandard.com	50 707

	right leaning news		center news		left leaning news	
	hostnames	N	hostnames	N	hostnames	N
1	wsj.com	310 416	cnn.com	2 291 736	nytimes.com	1 811 627
2	washingtontimes.com	208 061	thehill.com	1 200 123	washingtonpost.com	1 640 088
3	rt.com	157 474	politico.com	1 173 717	nbcnews.com	512 056
4	realclearpolitics.com	128 417	usatoday.com	326 198	abcnews.go.com	467 533
5	telegraph.co.uk	82 118	reuters.com	283 962	theguardian.com	439 580
6	forbes.com	64 186	bloomberg.com	266 662	vox.com	369 789
7	fortune.com	57 644	businessinsider.com	239 423	slate.com	279 438
8			apnews.com	198 140	buzzfeed.com	278 642
9			observer.com	128 043	cbsnews.com	232 889
10			fivethirtyeight.com	124 268	politifact.com	198 095
11			bbc.com	118 176	latimes.com	190 994
12			ibtimes.com	72 424	nydailynews.com	188 769
13			bbc.co.uk	71 941	theatlantic.com	177 637
14					mediaite.com	152 877
15					newsweek.com	149 490
16					npr.org	142 143
17					independent.co.uk	127 689
18					cnb.cx	87 094
19					hollywoodreporter.com	84 997

	left news		extreme bias (left) news	
	hostnames	N	hostnames	N
1	huffingtonpost.com	1 057 518	dailynewsbin.com	189 257
2	thedailybeast.com	378 931	bipartisanreport.com	119 857
3	dailykos.com	324 351	bluenationreview.com	75 455
4	rawstory.com	297 256	crooksandliars.com	73 615
5	politicususa.com	293 419	occupydemocrats.com	73 143
6	time.com	252 468	shareblue.com	50 880
7	motherjones.com	210 280	usuncut.com	27 653
8	talkingpointsmemo.com	199 346		
9	msnbc.com	177 090		
10	mashable.com	173 129		
11	salon.com	172 807		
12	thinkprogress.org	172 144		
13	newyorker.com	171 102		
14	mediamatters.org	152 160		
15	nymag.com	121 636		
16	theintercept.com	109 591		
17	thenation.com	54 661		
18	people.com	47 942		

Table A.1: **Hostnames in each media category in 2016**. We also show the number (N) of tweets with a URL pointing toward each hostname. Tweets with several URLs are counted multiple times. Reproduced from [18]

	fake news		extreme bias (right) news		right news	
	hostnames	N	hostnames	N	hostnames	N
1	thegatewaypundit.com	1 883 852	breitbart.com	2 192 997	foxnews.com	3 136 578
2	hannity.com	428 483	dailymail.co.uk	600 523	dailycaller.com	771 765
3	waynedupree.com	258 838	bongino.com	346 103	washingtonexaminer.com	717 017
4	judicialwatch.org	233 085	thenationalpulse.com	215 017	justthenews.com	689 725
5	truepundit.com	176 647	freebeacon.com	197 092	thefederalist.com	687 091
6	zerohedge.com	165 960	newsmax.com	192 924	dailywire.com	396 233
7	davidharrisjr.com	150 887	pjmedia.com	123 338	theepochtimes.com	288 656
8	politicalflare.com	145 838	newsbusters.org	71 008	nationalreview.com	283 172
9	djhjmedia.com	112 049	therightscoop.com	66 676	saraacarter.com	267 237
10	rumble.com	101 979	americanthinker.com	59 142	townhall.com	256 631
11	theconservativetreehouse.com	99 716			theblaze.com	191 515
12	oann.com	97 325			thepostmillennial.com	181 674
13	thedcpatriot.com	90 209			westernjournal.com	165 914
14	washingtonews.today	79 314			redstate.com	144 010
15	rightwingtribune.com	58 442			thegreggjarrett.com	139 749
16	rt.com	54 985			bizpacreview.com	97 375
17	wnd.com	54 929			twitchy.com	95 401
18	gellerreport.com	54 277			trendingpolitics.com	92 094
19	nationalfile.com	52 393			lifenews.com	90 064
20	summit.news	49 539				

	right leaning news		center news		left leaning news	
	hostnames	N	hostnames	N	hostnames	N
1	nypost.com	1 701 531	thehill.com	2 256 888	nytimes.com	6 775 402
2	wsj.com	887 537	apnews.com	1 182 504	washingtonpost.com	6 438 506
3	forbes.com	748 636	usatoday.com	993 957	cnn.com	5 577 352
4	washingtontimes.com	408 349	businessinsider.com	773 328	politico.com	2 290 755
5	foxbusiness.com	212 742	newsweek.com	756 820	nbcnews.com	2 231 564
6	thebulwark.com	175 417	reuters.com	746 033	theguardian.com	1 116 515
7	marketwatch.com	96 626	bbc.com	296 098	theatlantic.com	1 046 475
8	realclearpolitics.com	93 120	economist.com	123 939	abcnews.go.com	1 042 419
9	detroitnews.com	77 223	fivethirtyeight.com	101 824	npr.org	871 571
10	dallasnews.com	75 910	ft.com	91 524	bloomberg.com	767 059
11	rasmussenreports.com	58 712	foreignpolicy.com	87 729	cbsnews.com	747 442
12	chicagotribune.com	56 974	factcheck.org	79 456	cnbc.com	649 041
13	jpost.com	55 223	news.sky.com	78 372	axios.com	621 609
14					msn.com	613 127
15					news.yahoo.com	586 724
16					independent.co.uk	513 765
17					latimes.com	451 878
18					citizensforethics.org	382 101
19					buzzfeednews.com	369 962

	left news		extreme bias (left) news	
	hostnames	N	hostnames	N
1	rawstory.com	2 148 200	occupydemocrats.com	18 151
2	msnbc.com	1 606 071	lancastercourier.com	5815
3	thedailybeast.com	1 404 756	deepleftfield.info	5753
4	huffpost.com	1 121 642	tplnews.com	4022
5	politicususa.com	671 043	bipartisanreport.com	3243
6	palmerreport.com	434 503	bossip.com	2287
7	motherjones.com	424 106	polipace.com	586
8	vox.com	420 613		
9	vanityfair.com	352 964		
10	nymag.com	320 049		
11	newyorker.com	288 409		
12	dailykos.com	288 384		
13	slate.com	250 942		
14	salon.com	229 583		
15	rollingstone.com	190 828		
16	thenation.com	130 272		
17	alternet.org	126 788		
18	theintercept.com	104 153		

Table A.2: **Hostnames in each media category in 2020**. We also show the number (N) of tweets with a URL pointing toward each hostname. Tweets with several URLs are counted multiple times.

2016

	N_t	p_t	N_u	p_u	N_t/N_u	$p_{t,n/o}$	$p_{u,n/o}$	$N_{t,n/o}/N_{u,n/o}$
Fake news	2 991 073	0.10	68 391	0.03	43.73	0.19	0.07	124.22
Extreme bias right	3 969 639	0.13	131 346	0.06	30.22	0.09	0.05	56.73
Right news	4 032 284	0.13	194 229	0.08	20.76	0.11	0.07	33.77
Right leaning news	1 006 746	0.03	64 771	0.03	15.54	0.18	0.09	31.56
Center news	6 322 257	0.21	600 546	0.26	10.53	0.20	0.05	38.10
Left leaning news	7 491 344	0.24	903 689	0.39	8.29	0.14	0.06	19.16
Left news	4 353 999	0.14	327 411	0.14	13.30	0.14	0.07	26.16
Extreme bias left	609 503	0.02	19 423	0.01	31.38	0.06	0.03	74.21

2020

	N_t	p_t	N_u	p_u	N_t/N_u	$p_{t,n/o}$	$p_{u,n/o}$	$N_{t,n/o}/N_{u,n/o}$
Fake news	4 348 747	0.06	99 020	0.03	43.92	0.01	0.01	81.77
Extreme bias right	4 064 820	0.06	107 250	0.03	37.90	0.02	0.01	73.62
Right news	8 691 901	0.12	382 358	0.10	22.73	0.02	0.01	44.52
Right leaning news	4 648 000	0.06	288 207	0.08	16.13	0.02	0.01	23.35
Center news	7 568 472	0.10	398 241	0.11	19.00	0.03	0.02	33.96
Left leaning news	33 093 267	0.45	2 136 830	0.59	15.49	0.03	0.02	22.85
Left news	10 513 306	0.14	237 685	0.07	44.23	0.03	0.02	73.42
Extreme bias left	39 857	0.00	887	0.00	44.93	0.05	0.02	82.59

Table A.3: **Tweet and user volume corresponding to each media category on Twitter between June 1^{st} until election day in 2016 (top) and 2020 (bottom)**. Number, N_{t}, and proportion, p_{t}, of tweets with a URL pointing to a website belonging to one of the media categories. Number, N_{u}, and proportion, p_{u}, of unique users in each category. Users are classified in the category where the posted the largest number of tweets. Ties are randomly assigned. Proportion of tweets sent by non-official clients, $p_{\text{t,n/o}}$, proportion of users having sent at least one tweet from an non-official client, $p_{\text{u,n/o}}$, and average number of tweets per user sent from non-official clients, $N_{\text{t,n/o}}/N_{u,n/o}$.

	News category	Nodes	Edges	$\langle k \rangle$	max(k_{out})	max(k_{in})	$\sigma(k_{out})/\langle k \rangle$	$\sigma(k_{in})/\langle k \rangle$
2016	Fake News	175,605	1,143,083	6.51	42,468	1232	32 ± 4	2.49 ± 0.06
	Extreme bias (right)	249,659	1,637,927	6.56	51,845	588	36 ± 6	2.73 ± 0.03
	Right	345,644	1,797,023	5.20	86,454	490	44 ± 11	2.70 ± 0.04
	Right leaning	216,026	495,307	2.29	32,653	129	45 ± 11	1.72 ± 0.02
	Center	864,733	2,501,037	2.89	229,751	512	75 ± 39	2.69 ± 0.06
	Left leaning	1,043,436	3,570,653	3.42	145,047	843	59 ± 19	3.38 ± 0.10
	Left	536,903	1,801,658	3.36	58,901	733	47 ± 12	3.50 ± 0.08
	Extreme bias (left)	78,911	277,483	3.52	23,168	648	33 ± 6	2.49 ± 0.08
2020	Fake News	367,487	1,861,620	5.06	90,125	292	59 ± 11	2.05 ± 0.02
	Extreme bias (right)	445,776	2,008,760	4.50	89,902	313	60 ± 16	2.09 ± 0.02
	Right	674,935	4,452,861	6.59	109,053	607	54 ± 9	2.43 ± 0.03
	Right leaning	882,552	3,203,999	3.63	115,302	298	59 ± 16	1.86 ± 0.02
	Center	1,163,610	4,461,011	3.83	276,289	709	65 ± 29	2.37 ± 0.04
	Left leaning	2,355,587	17,461,102	7.41	325,726	1,564	63 ± 20	3.62 ± 0.05
	Left	819,684	4,688,119	5.71	175,841	1,042	57 ± 14	2.68 ± 0.04
	Extreme bias (left)	21,411	26,888	1.25	5,755	27	41 ± 3	0.60 ± 0.01

Table A.4: Retweet network characteristics for each news category. Number of nodes, edges, average degree and degree heterogeneity of each network. The in- and out-degree heterogeneities are calculated by taking the average and standard error of 1000 independent samples of the degree heterogeneity ($\sigma(k_{in})/\langle k \rangle$ and $\sigma(k_{out})/\langle k \rangle$), each of which is computed on 78,911 samples with replacements from their respective degree distributions.

rank	fake news (7 verified, 2 deleted, 19 unverified)	extreme bias (right) news (15 verified, 1 deleted, 9 unverified)	right news (22 verified, 0 deleted, 2 unverified)	right leaning news (20 verified, 1 deleted 4 unverified)
1	@PrisonPlanet ✓	@realDonaldTrump ✓	@FoxNews ✓	@WSJ ✓
2	@RealAlexJones ✓	@DailyCaller ✓	@realDonaldTrump ✓	@WashTimes ✓
3	@zerohedge	@BreitbartNews ✓	@dcexaminer ✓	@RT_com ✓
4	@DRUDGE_REPORT	@wikileaks ✓	@DRUDGE_REPORT	@realDonaldTrump ✓
5	@realDonaldTrump ✓	@DRUDGE_REPORT	@nypost ✓	@RT_America ✓
6	@mitchellvii ✓	@seanhannity ✓	@FoxNewsInsider ✓	@WSJPolitics ✓
7	deleted	@WayneDupreeShow ✓	@DailyMail ✓	@DRUDGE_REPORT
8	@TruthFeedNews	@LindaSuhler	@AllenWest ✓	@KellyannePolls ✓
9	@RickRWells	@mitchellvii ✓	@RealJamesWoods ✓	@TeamTrump ✓
10	deleted	@LouDobbs ✓	@foxandfriends ✓	@LouDobbs ✓
11	@gatewaypundit ✓	@PrisonPlanet ✓	@foxnation ✓	@rebeccaballhaus ✓
12	@infowars	@DonaldJTrumpJr ✓	@LouDobbs ✓	@WSJopinion ✓
13	@Lagartija_Nix	@gerfingerpoken	@KellyannePolls ✓	@reidepstein ✓
14	@DonaldJTrumpJr ✓	@FreeBeacon ✓	@JudicialWatch ✓	deleted
15	@ThePatriot143	@gerfingerpoken2	@PrisonPlanet ✓	@JasonMillerinDC ✓
16	@V_of_Europe	@TeamTrump ✓	@wikileaks ✓	@DanScavino ✓
17	@KitDaniels1776	@Italians4Trump	@TeamTrump ✓	@PaulManafort ✓
18	@Italians4Trump	@benshapiro ✓	@IngrahamAngle ✓	@SopanDeb ✓
19	@_Makada_	@KellyannePolls ✓	@marklevinshow ✓	@asamjulian
20	@BigStick2013	@DanScavino ✓	@LifeZette ✓	@JudicialWatch ✓
21	@conserv_tribune ✓	deleted	@theblaze ✓	@_Makada_
22	@Miami4Trump	@JohnFromCranber	@FoxBusiness ✓	@mtracey ✓
23	@MONAKatOILS	@true_pundit	@foxnewspolitics ✓	@Italians4Trump
24	@JayS2629	@ThePatriot143	@BIZPACReview	@Telegraph ✓
25	@ARnews1936	@RealJack	@DonaldJTrumpJr ✓	@RealClearNews ✓

rank	center news (24 verified, 0 deleted, 1 unverified)	left leaning news (25 verified, 0 deleted 0 unverified)	left news (21 verified, 0 deleted, 0 unverified)	extreme bias (left) news (7 verified, 1 deleted, 17 unverified)
1	@CNN ✓	@nytimes ✓	@HuffPost ✓	@Bipartisanism ✓
2	@thehill ✓	@washingtonpost ✓	@TIME ✓	@PalmerReport ✓
3	@politico ✓	@ABC ✓	@thedailybeast ✓	@peterdaou ✓
4	@CNNPolitics ✓	@NBCNews ✓	@RawStory ✓	@crooksandliars ✓
5	@Reuters ✓	@Slate ✓	@HuffPostPol ✓	@BoldBlueWave
6	@NateSilver538 ✓	@PolitiFact ✓	@NewYorker ✓	@Shareblue ✓
7	@AP ✓	@CBSNews ✓	@MotherJones ✓	@Karoli
8	@business ✓	@voxdotcom ✓	@TPM ✓	@RealMuckmaker
9	@USATODAY ✓	@ABCPolitics ✓	@Salon ✓	@GinsburgJobs
10	@AP_Politics ✓	@ezraklein ✓	@thinkprogress ✓	@AdamsFlaFan
11	@FiveThirtyEight ✓	@nytpolitics ✓	@mmfa ✓	@mcspocky
12	@bpolitics ✓	@guardian ✓	@joshtpm ✓	@Shakestweetz ✓
13	@jaketapper ✓	@NYDailyNews ✓	@MSNBC ✓	deleted
14	@DRUDGE_REPORT	@latimes ✓	@NYMag ✓	@JSavoly
15	@cnnbrk ✓	@BuzzFeedNews ✓	@samstein ✓	@OccupyDemocrats
16	@businessinsider ✓	@Mediaite ✓	@JuddLegum ✓	@ZaibatsuNews
17	@AC360 ✓	@HillaryClinton ✓	@mashable ✓	@wvjoe911
18	@cnni ✓	@nytopinion ✓	@theintercept ✓	@DebraMessing ✓
19	@brianstelter ✓	@CillizzaCNN ✓	@DavidCornDC ✓	@SayNoToGOP
20	@KellyannePolls ✓	@MSNBC ✓	@dailykos ✓	@coton_luver
21	@wikileaks ✓	@KFILE ✓	@JoyAnnReid ✓	@EJLandwehr
22	@SopanDeb ✓	@TheAtlantic ✓	@nxthompson ✓	@mch7576
23	@KFILE ✓	@SopanDeb ✓	@thenation ✓	@RVAwonk
24	@BBCWorld ✓	@Fahrenthold ✓	@justinjm1 ✓	@_Carja
25	@NewDay ✓	@BuzzFeed ✓	@ariannahuff ✓	@Brasilmagic

Table A.5: **Top 25 CI news spreaders of the retweet networks corresponding to each media category in 2016**. Verified users have a checkmark (✓) next to their username. Verifying its accounts is a feature offered by Twitter, that “lets people know that an account of public interest is authentic”. Unverified accounts do not have a checkmark and accounts marked as *deleted* have been deleted, either by Twitter or by the users themselves. Reproduced from [18].

rank	fake news (10 verified, 8 deleted, 7 unverified)	extreme bias (right) news (23 verified, 2 deleted, 0 unverified)	right news (23 verified, 1 deleted, 1 unverified)	right leaning news (23 verified, 2 deleted 0 unverified)
1	@seanhannity✓	(12) @DonaldJTrumpJr✓	(25) @DonaldJTrumpJr✓	@nypost✓
2	deleted	(3) @BreitbartNews✓	(19) @marklevinshow✓	(1) @WSJ✓
3	@DavidJHarrisJr	@dbongino✓	@jsolomonReports	@DonaldJTrumpJr✓
4	@JudicialWatch✓	@marklevinshow✓	(9) @RealJamesWoods✓	@EricTrump✓
5	@WayneDupreeShow✓	(1) @realDonaldTrump	(1) @FoxNews✓	(4) @realDonaldTrump
6	@catturd2	@newsmax✓	@SaraCarterDC✓	(2) @WashTimes✓
7	@TomFitton✓	@DailyMail✓	@DailyCaller✓	@marklevinshow✓
8	@OANN✓	@RaheemKassam✓	@MZHemingway✓	@brithume✓
9	@dbongino✓	@RealJamesWoods✓	@TrumpWarRoom✓	@RealJamesWoods✓
10	@Thomas1774Paine	@joelpollak✓	(3) @dcexaminer✓	@KimStrassel✓
11	@RealMattCouch	@JackPosobiec✓	@JackPosobiec✓	@newtgingrich✓
12	deleted	@TomFitton✓	@seanmdav✓	@TrumpWarRoom✓
13	(3) @zerohedge	@TrumpWarRoom✓	@realDailyWire✓	deleted
14	@Rasmussen_Poll✓	@RCamposDuffy✓	@GOPChairwoman✓	@MichaelCBender✓
15	@atensnut	@EricTrump✓	(2) @realDonaldTrump	@RandPaul✓
16	(1) @PrisonPlanet✓	@JasonMillerinDC✓	@GreggJarrett✓	(15) @JasonMillerinDC✓
17	@CassandraRules✓	(14) @FreeBeacon✓	@newtgingrich✓	@JackPosobiec✓
18	deleted	@AlexMarlow✓	@kayleighmcenany✓	@BillKristol✓
19	@DineshDSouza✓	@bennyjohnson✓	@RepDougCollins✓	@AriFleischer✓
20	(5) @realDonaldTrump	@FrankelJeremy✓	@RichardGrenell✓	@Rasmussen_Poll✓
21	@HowleyReporter	deleted	@AndrewCMcCarthy✓	@IngrahamAngle✓
22	deleted	@SteveGuest✓	@SteveGuest✓	@RudyGiuliani✓
23	deleted	@BrentScher✓	@SecretsBedard✓	@MZHemingway✓
24	deleted	@IngrahamAngle✓	@parscale✓	@Forbes✓
25	deleted	@kimguilfoyle✓	@dbongino✓	(11) @rebeccaballhaus✓
rank	center news (24 verified, 0 deleted, 1 unverified)	left leaning news (24 verified, 0 deleted 1 unverified)	left news (23 verified, 0 deleted, 2 unverified)	extreme bias (left) news (3 verified, 1 deleted, 21 unverified)
1	(2) @thehill✓	@CNN✓	(13) @MSNBC✓	@DearAuntCrabby
2	(7) @AP✓	(1) @nytimes✓	(3) @thedailybeast✓	@funder✓
3	(5) @Reuters✓	@kylegriffin1✓	@kylegriffin1✓	@ImpeachmentHour
4	@kylegriffin1✓	(3) @ABC✓	(19) @DavidCornDC✓	@MeidasTouch
5	@JonLemire✓	(2) @washingtonpost✓	(1) @HuffPost✓	@TheDemCoalition✓
6	@Newsweek✓	@CNNPolitics✓	@NoahShachtman✓	@grantstern✓
7	@yarotrof✓	@NPR✓	(4) @RawStory✓	(15) @OccupyDemocrats
8	(9) @USATODAY✓	(4) @NBCNews✓	(7) @MotherJones✓	@Stop_Trump20
9	@ProjectLincoln	(7) @CBSNews✓	@TeaPainUSA	@InSpiteOfTrump
10	@JoeBiden✓	@politico✓	@svdate✓	@froggneal
11	@TheDemCoalition✓	@ddale8✓	@voxdotcom✓	@atav1k
12	@TheEconomist✓	@CREWcrew✓	@maddow✓	@diamondlilron
13	(10) @AP_Politics✓	@cnnbrk✓	@joncoopertweets✓	@HollyHuntley3
14	@TheRickWilson✓	@maddow✓	@Slate✓	deleted
15	@tribelaw✓	@jaketapper✓	@PoliticusSarah✓	@patrickinmass
16	@SkyNews✓	@ThePlumLineGS✓	@tribelaw✓	@Franpianos
17	@maddow✓	@NatashaBertrand✓	@JoeBiden✓	@willapercy
18	@FinancialTimes✓	@tribelaw✓	@TheRickWilson✓	@Jerrygence
19	@joncoopertweets✓	@axios✓	@realTuckFrumper	@bethlevin
20	@FrankFigliuzzi1✓	(18) @nytopinion✓	@VanityFair✓	@nyx_with
21	@JimLaPorta✓	@maggieNYT✓	@CREWcrew✓	@vg123e
22	@DonaldJTrumpJr✓	(14) @latimes✓	(6) @NewYorker✓	@watercutter11
23	(24) @BBCWorld✓	@ProjectLincoln	@PalmerReport✓	@404HDTV
24	@APFactCheck✓	@60Minutes✓	@11thHour✓	@j_starace
25	@KamalaHarris✓	@business✓	(5) @HuffPostPol✓	@amberofmanyhats

Table A.6: **Top 25 CI news spreaders of the retweet networks corresponding to each media category in 2020**. Verified users have a checkmark (✓) next to their username. Unverified accounts do not have a checkmark and accounts marked as *deleted* have been deleted, either by Twitter or by the users themselves. If a user held a position in the top 25 in 2016 as well, we mark that position for reference in parentheses next to the username. Despite *@realDonaldTrump* having their account permanently suspended, due to the role they played in the 2020 Election, we have chosen to keep their original Twitter username in the table. However, we count this account as deleted, and have removed their previously assigned checkmark.

Year	Modularity (SE)	Normalized Cut (SE)	Right Ratio	Left Ratio
2016	0.234 (0.004)	0.66 (0.03)	0.038	0.05
2020	0.236 (0.007)	0.58 (0.03)	0.038	0.08

Table A.7: Tabulated analysis of the similarity network using quotes instead of retweets for the top influencers (as determined by the CI rankings of the retweet networks). The similarity network is found for the 2016 and 2020 data. Using Louvain community detection reveals two communities with left- and center-oriented influencers in one community, and right- and fake-oriented influencers in the other. Left side of table: average modularity and average normalized cut, with the standard errors (SE) in parentheses, determined by taking sub-samples of influencers from the quote similarity network, detecting the two dichotomous communities with the sub-sampled quote similarity network, then recording their modularities and normalized cuts. Right side of table: ratio of quotes-to-retweets within the complete similarity network. Specifically, number of user quotes of influencer tweets over number of user retweets of influencer tweets. Right ratio indicates the average ratio for the community with right-oriented influencers. Left ratio indicates the average ratio for the community with left-oriented influencers. These ratios are found for both 2016 and 2020.

A overall quotes/retweets

		2016	2020
from users	right	0.03	0.03
	left	0.05	0.04

B quotes/retweets

		2016		2020	
		to influencers			
		right	left	right	left
from users	right	0.02	0.19	0.02	0.49
	left	0.56	0.03	3.76	0.03

Table A.8: Comparison of fraction of retweets and quotes from users to influencers with different latent ideology estimates. Users and influencers are divided in two categories based on their ideology estimates, namely left (ideology <0) and right (ideology>0). Table **A** shows the overall proportion of quotes over retweets from users on the right and on the left revealing that the number of quotes represent only a small fraction ($\leq 5\%$) of the number of retweets. Table **B** shows the proportion of quotes over retweets from users to influencers for all pairs of ideology categories in 2016 and in 2020.

	users distributions					influencers distributions				
	2016	95% CI	2020	95% CI	difference	2016	95% CI	2020	95% CI	difference
all	0.1086	[0.1082,0.1091]	0.1474	[0.1471,0.1477]	0.0388	0.1786	[0.1606,0.1965]	0.2091	[0.1907,0.2282	0.0305
common users	0.0941	[0.0934,0.0947]	0.1172	[0.1166,0.1178]	0.0231	0.1793	[0.1616,0.1979]	0.2143	[0.1952,0.2336]	0.0350
common influencers	0.1070	[0.1065,0.1076]	0.1830	[0.1825,0.1834]	0.0760	0.1641	[0.1290,0.1951]	0.1741	[0.1376,0.2122]	0.0100
common users and influencers	0.0947	[0.0940,0.0955]	0.1399	[0.1390,0.1406]	0.0452	0.1650	[0.1314,0.2034]	0.1719	[0.1379,0.2086]	0.0069

Table A.9: Hartigans' dip test statistics of the users and influencers latent ideology distributions when considering all users and influencers, only users that were present in 2016 and 2020, only influencers that were present in 2016 and 2020 and only users and influencers that were present in 2016 and 2020. 95% confidence intervals are computed from 1000 bootstrap samples with the bias-corrected and accelerated confidence intervals method.