ISTITUTO DI INFORMATICA E TELEMATICA
Consiglio Nazionale delle Ricerche

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | SOBIGDATA.it ITALIAN RESEARCH INFRASTRUCTURE | Consiglio Nazionale delle Ricerche

Oggetto: Trasmissione tracce prova orale relative alla Selezione per titoli e colloquio ai sensi dell'art. 8 del *"Disciplinare concernente le assunzioni di personale con contratto di lavoro a tempo determinato"*, per l'assunzione, ai sensi dell'art. 83 del CCNL del Comparto "Istruzione e Ricerca" 2016-2018, sottoscritto in data 19 aprile 2018, di una unità di personale con profilo professionale di un **Ricercatore III livello**, presso l'Istituto di Informatica e Telematica (IIT)- sede Pisa
BANDO N. 400.01 IIT PNRR

In relazione al bando in oggetto si dispone la pubblicazione sulla pagina del sito Internet del CNR agli indirizzi https://www.urp.cnr.it/ e https://www.selezionionline.cnr.it/ delle buste contenenti le domande delle prove orali allegate al presente provvedimento.

La responsabile del procedimento
Dott.ssa Irene Sannicandro

BUSTA 1

1. Il candidato descriva le maggiori criticità a cui si va incontro quando si raccolgono dati da social media
2. Il candidato esponga le differenze tra Deep Learning e Machine Learning
3. Il candidato descriva alcune tecniche per fare Cyber Intelligence utilizzando dati proveniente da Social Media

INGLESE: Abstract del paper: Flamino, J., Galeazzi, A., Feldman, S. *et al.* Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections. *Nat Hum Behav* (2023). https://doi.org/10.1038/s41562-023-01550-8 𝒶 ℰ 𝒬

nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01550-8

Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections

Received: 28 January 2022

Accepted: 3 February 2023

Published online: 13 March 2023

Check for updates

James Flamino[1], Alessandro Galeazzi[2,3], Stuart Feldman[4], Michael W. Macy[5], Brendan Cross[1], Zhenkun Zhou[6], Matteo Serafino[7], Alexandre Bovet[8], Hernán A. Makse[7] & Boleslaw K. Szymanski[1]

busta 1

Social media has been transforming political communication dynamics for over a decade. Here using nearly a billion tweets, we analyse the change in Twitter's news media landscape between the 2016 and 2020 US presidential elections. Using political bias and fact-checking tools, we measure the volume of politically biased content and the number of users propagating such information. We then identify influencers—users with the greatest ability to spread news in the Twitter network. We observe that the fraction of fake and extremely biased content declined between 2016 and 2020. However, results show increasing echo chamber behaviours and latent ideological polarization across the two elections at the user and influencer levels.

busta 2

A growing number of studies have documented increasing political polarization in the USA that is deeper than at any time since the American Civil War[1–3]. 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[4,5]. 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 voting data that lack relational measures. Increasing reliance on social media for political communication is opening unprecedented opportunities to study the diffusion of political information and misinformation[6,7] over communication networks[8]. Furthermore, 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 can 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[9–18], including the effects of disinformation[19–25]. Other studies have documented how social media platforms contribute to polarization through the creation of echo chambers[26–35].

We use a vast amount of social media data collected from Twitter over the 2016 and 2020 US presidential elections enriched with political bias classifications to study diffusion dynamics of political content through news media. In this longitudinal study, we focus on shifts in Twitter's political landscape caused by changes in the news media content being disseminated. We discovered that, proportionally, the fraction of tweets in the fake news and extremely biased news categories decreased or stayed the same on Twitter.

We also focus on analysing news media influencers, defined as users with the greatest ability to broadly propagate news media information over social media. We analyse changes in their influence, composition and the types of news media they are disseminating between

[1]Department of Computer Science and Network Science and Technology Center, Rensselaer Polytechnic Institute, Troy, NY, USA. [2]University of Brescia, Brescia, Italy. [3]Ca' Foscari University of Venice, Venice, Italy. [4]Schmidt Futures, New York, NY, USA. [5]Departments of Information Science and Sociology, Cornell University, Ithaca, NY, USA. [6]School of Statistics, Capital University of Economics and Business, Beijing, China. [7]Levich Institute and Physics Department, City College of New York, New York, NY, USA. [8]Department of Mathematics and Digital Society Initiative, University of Zurich, Zurich, Switzerland. e-mail: hmakse@ccny.cuny.edu; szymab@rpi.edu

BUSTA 2

1. Il candidato descriva la differenza tra web scraper e web crawler
2. Il candidato descriva i limiti del Deep Learning
3. Il candidato illustri un sistema di storage per Big Data che può essere utilizzato in ambito Cyber Intelligence

INGLESE: primi paragrafi (in giallo) del paper: Flamino, J., Galeazzi, A., Feldman, S. *et al.* Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections. *Nat Hum Behav* (2023). https://doi.org/10.1038/s41562-023-01550-8

nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01550-8

Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections

Received: 28 January 2022

Accepted: 3 February 2023

Published online: 13 March 2023

Check for updates

James Flamino[1], Alessandro Galeazzi[2,3], Stuart Feldman[4], Michael W. Macy[5], Brendan Cross[1], Zhenkun Zhou[6], Matteo Serafino[7], Alexandre Bovet[8], Hernán A. Makse[7] & Boleslaw K. Szymanski[1]

busta 1

Social media has been transforming political communication dynamics for over a decade. Here using nearly a billion tweets, we analyse the change in Twitter's news media landscape between the 2016 and 2020 US presidential elections. Using political bias and fact-checking tools, we measure the volume of politically biased content and the number of users propagating such information. We then identify influencers—users with the greatest ability to spread news in the Twitter network. We observe that the fraction of fake and extremely biased content declined between 2016 and 2020. However, results show increasing echo chamber behaviours and latent ideological polarization across the two elections at the user and influencer levels.

busta 2

A growing number of studies have documented increasing political polarization in the USA that is deeper than at any time since the American Civil War[1–3]. 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[4,5]. 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 voting data that lack relational measures. Increasing reliance on social media for political communication is opening unprecedented opportunities to study the diffusion of political information and misinformation[6,7] over communication networks[8]. Furthermore, 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 can 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[9–18], including the effects of disinformation[19–25]. Other studies have documented how social media platforms contribute to polarization through the creation of echo chambers[26–35].

We use a vast amount of social media data collected from Twitter over the 2016 and 2020 US presidential elections enriched with political bias classifications to study diffusion dynamics of political content through news media. In this longitudinal study, we focus on shifts in Twitter's political landscape caused by changes in the news media content being disseminated. We discovered that, proportionally, the fraction of tweets in the fake news and extremely biased news categories decreased or stayed the same on Twitter.

We also focus on analysing news media influencers, defined as users with the greatest ability to broadly propagate news media information over social media. We analyse changes in their influence, composition and the types of news media they are disseminating between

[1]Department of Computer Science and Network Science and Technology Center, Rensselaer Polytechnic Institute, Troy, NY, USA. [2]University of Brescia, Brescia, Italy. [3]Ca' Foscari University of Venice, Venice, Italy. [4]Schmidt Futures, New York, NY, USA. [5]Departments of Information Science and Sociology, Cornell University, Ithaca, NY, USA. [6]School of Statistics, Capital University of Economics and Business, Beijing, China. [7]Levich Institute and Physics Department, City College of New York, New York, NY, USA. [8]Department of Mathematics and Digital Society Initiative, University of Zurich, Zurich, Switzerland. e-mail: hmakse@ccny.cuny.edu; szymab@rpi.edu

the two elections. We find that the proportion of top influencers affiliated with news media organizations decreased in 2020, while the proportion of those affiliated with political organizations increased. We also quantify and compare the levels of polarization between 2016 and 2020. There are multiple types and levels of polarization established in literature[36–43], which we discuss in the Supplementary Materials. However, we focus on 'ideological polarization'[44] of Twitter users, defined as the level of ideological separation between the political alignments of the content that these users propagate. For the remainder of the article, we use the term 'polarization' to refer specifically to 'ideological polarization'. Our polarization analysis reveals an increase in echo chamber behaviour between 2016 and 2020 resulting from Twitter users' tendency to be less likely to disseminate information or interact with users on the other side of the political spectrum. This analysis also suggests that new influencers from 2020 are more polarized than the influencers who persisted from the 2016 US presidential election. We believe these results establish a foundation for future work by providing observations on trends and patterns arising in Twitter's political landscape in news media.

Results

We note that the initial foundation for this research is established in ref. [21], which analysed the news media diffusion dynamics on Twitter during the 2016 US presidential election. We harness part of the data used in that article and follow its relevant methodology to identify and classify influencers in the 2020 US election data. Additionally, following an editorial request added to the reviews of this article, we anonymized all Twitter usernames of personal accounts in both the main manuscript and the Supplementary Materials. Specifically, if the username being presented does not represent an established major news organization that is verified on Twitter, that username is replaced with an alias. This alias consists of two parts: affiliation and year of relevance. A user's affiliation can be with the media, US politics or personal (see the News media influencers section for more information on how we define affiliations). The personal affiliation is also split into 'individual' and 'other' labels, with the former representing no official affiliation with media or politics, and the latter representing a lack of information required to make a distinction. All affiliation labels are shortened to their first five letters in the alias. Year of relevance is determined as being in the top 100 list of influencers for 2016, 2020 or both. See the Twitter retweet networks section for more details on influencers and our influencer identification algorithm. So, a politically affiliated user that was influential only in 2016 will have an alias of 'Polit_2016'.

News media on Twitter in 2016 and 2020

We tracked the spread of political news on Twitter in 2016 and 2020 by analysing two datasets containing tweets posted between 1 June and election day (8 November in 2016 and 2 November 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. Using more keywords targeting specific media outlets or hashtags concerning specific news events could miss 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. [13,21] 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 near doubling of the number of Twitter users involved in spreading political news in 2020 compared with 2016.

At the time we collected our data, the statistical analyses of the raw collected data were limited because the data collection process designed by Twitter itself has been shown to have sampling issues. For instance, the probability of non-responses from API queries is not provided by Twitter, and Twitter has acknowledged that the 100% firehose is not actually a 100% sample, the 10% is not a randomly distributed 10% and the 1% is not a randomly distributed 1%. Thus, standard sampling methods are difficult to apply to the collected Twitter data. However, for the goals of our article, this is our best option as there are no other large-scale, comprehensive datasets available for both the 2016 and 2020 US elections that are readily accessible to us.

The classifications of news media websites presented below and used here, including 'fake', 'extremely biased', 'left' and 'right', and especially the boundaries between categories, are a matter of opinion rather than a statement of fact. We use terms 'left' and 'right' for political leanings that are often referred to as 'liberal' and 'conservative' on the US political ideology spectrum. The categorizations and labels assigned to the corresponding classes and used here originated in publicly available datasets from fact-checking and bias rating organizations, which are credited below. The classifications of political views and the related conclusions contained in this article should not be interpreted as representing opinions of the authors or their funders.

For each tweet containing a URL link, we extracted the domain name of the URL (for example, 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 the website mediabiasfactcheck.com (MBFC) for outlets not listed in AS (both accessed on 7 January 2021 for the 2020 classification). We classified URL links for outlets that mostly conform to professional standards of fact-based journalism in five news media categories: right, right leaning, centre, left leaning and left. We also include three additional news media categories to include outlets that tend to disseminate disinformation: extreme bias right, extreme bias left 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. The full lists of outlets in each category in 2016 and 2020 are given in Supplementary Tables 1 and 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 a drop in the fraction of tweets flowing from users that propagate news media links, from 18% in 2016 to 10% in 2020.

The proportions of tweets and users who sent a tweet in each of the news media categories are shown in Fig. 1a,b along with other statistics about the activity of users in each category. The raw numbers used to generate this figure are shown in Supplementary Table 3. Importantly, they demonstrate that 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 extreme bias right news. The fraction of users who shared those tweets also decreased for extreme bias right news (from 6 to 3%) but not for fake news (which remained at 3%). However, the total number of tweets and users increased over the same period by 411 and 80%, respectively. In short, between 2016 and 2020, the numbers of tweets and users grew at a rate in the range of 80 to 246% for all categories, except the number of users who shared extreme bias right news, which declined by 10%.

The fraction of tweets in the extreme bias left category was only 2% in 2016 and it dropped to a mere 0.05% in 2020. The number of tweets in this category also dropped. The fraction of tweets in the centre category also decreased, from 21 to 10%, but the number of tweets increased dramatically. By contrast, the fraction of left-leaning tweets increased from 24 to 45%, while the fraction of right-leaning tweets increased from 3 to 6%.