Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Information disorders during the COVID-19 infodemic: The case of Italian Facebook

Stefano Guarino a,∗, Francesco Pierri a,∗, Marco Di Giovanni b, Alessandro Celestini a

a Institute for Applied Mathematics, National Research Council, 00185, Rome, Italy
b Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133, Milan, Italy

ARTICLE INFO

Keywords:
COVID-19
Disinformation
Infodemic
Online social networks
Facebook

ABSTRACT

The recent COVID-19 pandemic came alongside with an “infodemic”, with online social media flooded by often unreliable information associating the medical emergency with popular subjects of disinformation. In Italy, one of the first European countries suffering a rise in new cases and dealing with a total lockdown, controversial topics such as migrant flows and the 5G technology were often associated online with the origin and diffusion of the virus. In this work we analyze COVID-19 related conversations on the Italian Facebook, collecting over 1.5M posts shared by nearly 80k public pages and groups for a period of four months since January 2020. On the one hand, our findings suggest that well-known unreliable sources had a limited exposure, and that discussions over controversial topics did not spark a comparable engagement with respect to institutional and scientific communication. On the other hand, however, we realize that dis- and counter-information induced a polarization of (clusters of) groups and pages, wherein conversations were characterized by a topical lexicon, by a great diffusion of user generated content, and by link-sharing patterns that seem ascribable to coordinated propaganda. As revealed by the URL-sharing diffusion network showing a “small-world” effect, users were easily exposed to harmful propaganda as well as to verified information on the virus, exalting the role of public figures and mainstream media, as well as of Facebook groups, in shaping the public opinion.

1. Introduction and related work

The spread of a novel coronavirus (SARS-CoV-2) in the past months has changed in an unprecedented way the everyday life of people on a global scale. According to the World Health Organization (WHO), at the time of this writing the pandemic has caused over 23M confirmed cases, with more than 809k fatalities globally speaking.1 Italy, in particular, has been one of the first European countries to be severely hit by the pandemic, as the virus spread outside China borders at the end of January, and to implement national lockdown on the 8th of March [1,2]. Following Italy and China, national lockdowns have been adopted by most countries around the world, drastically reducing mobility flows in order to circumvent the spread [3].

In relation to the emergency, the term “infodemic” has been coined to describe the risks related to the massive spread of harmful and malicious content on online social platforms [4], as misinformation could support the spread of the virus undermining medical efforts and, at the same time, drive societal mistrust producing other direct damages [4]. In response, several contemporary works have provided different perspectives on this phenomenon. Authors of [5] analyzed more than 100 millions Twitter messages posted worldwide in 64 languages and found correspondence between waves of unreliable and low-quality information and the epidemic ones. Authors of [6] have investigated the prevalence of low-credibility content in relation to the activity of social bots, showing that the combined amount of unreliable information is comparable to the retweets of articles published on The New York Times alone. Finally, authors of [7] carried out a comparative analysis of information diffusion on different social platforms, from Twitter to Reddit, finding different volumes of misinformation in different environments.

As a matter of fact, ever since 2016 US presidential elections we observed a growing concern of the research community over deceptive information spreading on social networks [8–11]. In Italy, according to Reuters, trust in news is particularly low today [12], and previous research has highlighted the exposure to online disinformation in several political circumstances, from 2016 Constitutional Referendum to 2019 European Parliament elections [13–17]. A recent questionnaire by the EU funded SOMA observatory on disinformation spreading on online social media showed that people relied on official

https://doi.org/10.1016/j.osnem.2021.100124

Received 30 August 2020; Received in revised form 30 December 2020; Accepted 23 January 2021
Available online 8 February 2021
2468-6964/© 2021 Elsevier B.V. All rights reserved.
channels used by authoritative institutions in order to inform about the pandemic. Interestingly, social media were not the primary source of information during the crisis.

Similar to contemporary research, in this work we adopt a consolidated strategy to label news articles at the source level \cite{10,18–21} and investigate accordingly the diffusion of different kinds of information on Facebook. Thus, we use the term “disinformation” as a shorthand for unreliable information in several forms, all potentially harmful, including false news, click-bait, propaganda, conspiracy theories and unverified rumours. We use instead the term “mainstream” to indicate traditional news websites which convey reliable and accurate information. This approach has been mainly used for Twitter, which however exhibits a declining trend as a platform to consume online news \cite{12,17}. Similar to \cite{22}, we leverage CrowdTangle platform to collect posts related to COVID-19 from Facebook public pages and groups. We use a set of keywords related to the epidemic and we limit the search to posts in the Italian language. The overall dataset accounts for over 1.5M public posts shared by almost 80k unique pages/groups.

We investigate the prevalence of reliable vs non-reliable information by analyzing the domain of URLs included in such posts. In particular, we are interested in understanding how specific disinformation narratives compete with official communications. To this aim, we further specify keywords related to three different controversial topics that have been trending in the past few months, all related to the origins of the novel coronavirus: (1) the alleged correlation between COVID-19 and migrants, (2) between the virus and 5G technology, and (3) rumours about the artificial origin of the virus.

This work provides the following main contributions:

- We evaluate the prevalence of COVID-19 and related controversial topics on the Italian Facebook, identifying the key players and the most relevant pieces of content in the information ecosystem in terms of both volume of posts and generated engagement.
- We study how these issues shaped the debate on Facebook, quantifying the sentiment of posts, the polarization of groups and pages w.r.t. topics of discussion and measuring the respective lexical/semantic divergence.
- We analyze patterns in the URL sharing network of groups/pages, observing that the majority of Facebook groups and pages interact in a “small-world” – thus discarding the hypothesis that different groups draw upon fully separated pools of web resources.
- We focus on the connections among URLs related to controversial topics and among groups/pages where these URLs were shared, to find evidence of a coordinated effort to spread propaganda and to ascertain that centrality in these networks is not directly correlated with high engagement.

The outline of this paper is the following: we first describe the methodology applied, including the collection of data from Facebook, the taxonomy of news sources and controversial topics, and both text and network analysis tools (Section 2); then we describe our contributions (Sections 3 and 4), and finally we draw conclusions and future work (Section 5).

2. Methodology

2.1. Facebook data collection

We used CrowdTangle’s “historical data” interface \cite{23} to fetch posts in Italian language shared by public pages and groups since January 1st 2020 until May 12th 2020 and containing any of the following keywords: virus, coronavirus, covid, sars-cov-2, sars cov 2, pandemic, epidemic, pandemic, epidemic. The tool only tracks public posts made by public accounts or groups. Besides, it does not track every public account \footnote{All pages with at least 100 K likes are fully retained. For details on the coverage for pages with less likes we refer the reader to \url{https://help.crowdtangle.com/en/articles/1140930-what-is-crowdtangle-tracking}.} and does not track neither private profiles nor private groups. For each post we collected the number of public interactions, namely, likes, reactions, comments, shares, upvotes and three second views, as well as Uniform Resource Locators (URL) attached with it. Our collection contains overall 1.59M posts shared by 87,426 unique Facebook pages/groups. In the rest of the paper, we use “accounts” as a shorthand to indicate the entire set of pages and groups. Data is not publicly available, but it can be provided to academics and non-profit organizations upon request to the platform.

2.2. Mainstream and disinformation news

Similarly to \cite{5–7}, we aim to understand the prevalence of reliable vs non-reliable information based on lists of news outlets compiled from multiple sources and employed in a previous analysis of the Italian information ecosystem \cite{17,21}. We use a coarse “source-based” approach, widely adopted in the literature \cite{10,18–20}, to label links shared in Facebook posts that point to online news articles in two classes according to their domain: (i) Disinformation sources, which notably publish a variety of harmful information, from hyper-partisan stories to false news and conspiracy theories; (ii) Mainstream sources, which may be assumed to generally provide accurate and reliable news reporting. Indeed, this classification might not always hold since unreliable websites do share also true news, and incorrect news coverage on traditional outlets is not rare \cite{9}; however, it has been proven effective to analyze content shared during the 2016 US Presidential elections \cite{10,18,19}. For what concerns unreliable news, we further partition the class into four distinct sets according to the geographic area: European (EU), Italian (IT), Russian (RU) and US sources. The overall list, available in the Appendix, Tables A.1–A.5, contains 25 Italian sources for the Mainstream domain whereas, for the disinformation domain, we count 25 EU sources, 52 Italian sources, 13 Russian sources and 22 US sources.

2.3. Controversial topics

In our analysis, we focus on three specific topics which were particularly exposed to disinformation during the infodemic \footnote{\url{https://www.newsguardtech.com/covid-19-myths/}}:

- **Migrants**: conspiracy theories that attempt to correlate the spread of the virus with migration flows. These are mainly promoted by far-right communities to foster racial hate. Some of the related keywords are: migrants, immigranti, ong, barconi, extracomunitari, africa.
- **LABS**: rumours that have been used as political weapons to attribute the origins of the pandemic to the development of a bioweapon to be used by China and/or to undermine the forthcoming U.S. presidential elections. Some of the related keywords are: laboratorio, ricerca, sperimentazione.
- **5G**: hoaxes that can be summarized in two main streams, those claiming that 5G activates COVID-19 and those that deny the existence of the novel coronavirus and attribute its symptoms to reactions to 5G waves. Both lines are obviously false and not supported by scientific evidence. Some of the related keywords are: 5g, onde, radiazioni, elettromagnetismo.

A complete list of keywords for each topic is available in the Appendix. For the sake of simplicity, we will refer to an account as a “Migrants” account – and likewise for the other topics – if the account shared at least \( N = 2 \) posts (to reduce noise) which contain a keyword in the related list; the same holds for URLs if the associated post contained...
original bipartite graph. This means that two URLs to identify statistically significant patterns of common neighbors in the
tions, relying on the Bipartite Configuration Model (BiCM)
projected graph. We then focus on two statistically validated projec-
to identify statistically significant patterns of common neighbors in the
bipartite graph is sufficient for being connected in the
We first implement two naive projections, wherein having one common
neighbor in the bipartite graph is sufficient for being connected in the
We leverage tools from network science \[26,27\] in order to inves-
tigate the diffusion network of content shared on Facebook. Similar
to \[22\], we use a bipartite graph formulation to link together accounts
and URLs. Precisely, we draw an undirected edge between an account
\( a \) and an URL \( u \) if and only if \( u \) was shared at least once on/by \( a \). This graph has 983,582 vertices (78,760 accounts and 904,822 URLs)
and 1,374,921 edges. We then focus on the controversial bipartite
graph defined as the subgraph of the accounts-URLs graph induced
by controversial URLs. This subgraph has 55,411 vertices (18,681
accounts and 1,374,921 edges. We then focus on the controversial bipartite
This graph has 983,582 vertices (78,760 accounts and 904,822 URLs)
accounts \( a_1 \), \( u_1 \) and \( a_2 \) are connected if the number of accounts that shared
both (resp., of URLs shared by both) is too large to be just explained
by the degree distribution of the two layers.

2.5. Network analysis

We clean and pre-process posts’ textual content as follows, relying
on the spaCy \[24\] and nltk \[25\] Python libraries. Firstly, we lower-
case all strings and we remove URLs, punctuation, emojis and Italian
stop words. We also remove words related to the COVID-19 as they act
as stop word for our analysis. Then, we tokenize texts and we remove
tokens shorter than 4 or longer than 20 characters.

Then, we group tokens by account. To reduce noise effects we
remove accounts with only 1 post and accounts with less than 20
tokens in total, obtaining 56,436 accounts from an original amount of
87,426. We compute the TF-Idf of the cleaned strings, neglecting tokens
that appeared less than 5 times in the whole corpus. Finally, for each
account we obtain a sparse 137,901-dimensional embedding vector.

2.4. Text analysis

We clean and pre-process posts’ textual content as follows, relying
on the spaCy \[24\] and nltk \[25\] Python libraries. Firstly, we lower-
case all strings and we remove URLs, punctuation, emojis and Italian
stop words. We also remove words related to the COVID-19 as they act
as stop word for our analysis. Then, we tokenize texts and we remove
tokens shorter than 4 or longer than 20 characters.

Then, we group tokens by account. To reduce noise effects we
remove accounts with only 1 post and accounts with less than 20
tokens in total, obtaining 56,436 accounts from an original amount of
87,426. We compute the TF-Idf of the cleaned strings, neglecting tokens
that appeared less than 5 times in the whole corpus. Finally, for each
account we obtain a sparse 137,901-dimensional embedding vector.

2.5. Network analysis

We leverage tools from network science \[26,27\] in order to inves-
tigate the diffusion network of content shared on Facebook. Similar
to \[22\], we use a bipartite graph formulation to link together accounts
and URLs. Precisely, we draw an undirected edge between an account
\( a \) and an URL \( u \) if and only if \( u \) was shared at least once on/by \( a \). This graph has 983,582 vertices (78,760 accounts and 904,822 URLs)
and 1,374,921 edges. We then focus on the controversial bipartite
graph defined as the subgraph of the accounts-URLs graph induced
by controversial URLs. This subgraph has 55,411 vertices (18,681
accounts and 1,374,921 edges. We then focus on the controversial bipartite
This graph has 983,582 vertices (78,760 accounts and 904,822 URLs)
accounts \( a_1 \), \( u_1 \) and \( a_2 \) are connected if the number of accounts that shared
both (resp., of URLs shared by both) is too large to be just explained
by the degree distribution of the two layers.

3. Descriptive statistics

3.1. Posts, interactions and news articles

We first inspect the prevalence of COVID-19 in online conversations
by looking at the time series of daily posts on Facebook, reported in
the left panel of Fig. 1 for both the whole dataset (bottom) and the
three controversial topics (top). We observe a general increase in the
overall volume, with a few notable spikes: at the end of January, when
China imposed the lockdown; at the end of February, when the virus
was first diagnosed in Italy; at mid March, when the lockdown was
applied in Italy; at the beginning of May, when the restrictions have
been lifted. For what concerns controversial topics, we immediately see
that the volumes are negligible w.r.t general conversations and that
the trends are quite aligned. Yet, a few topic specific spikes can be observed, which are most likely related to real world events such as the
sabotages of 5G antennas in several countries. In the Appendix, Fig. A.1,
we provide the time series of the daily engagement, which is by all
means analogous to the volume of posts, although with a different order
of magnitude — the total daily engagement reaches \(10^{19}\), while the
engagement of controversial topics is consistently smaller by 2 orders
of magnitude.

For what concerns the diffusion of URLs, we inspect most popular
domains by focusing on their total engagement. In particular, in the
Top-10 ranking of domains we encounter websites which are all related
to Italian Mainstream newspapers, with the exception of Facebook and
YouTube which are the 2 most shared domains. When we focus on the
Top-10 ranking of news websites, we observe what follows (see also
Fig. A.2 in the Appendix):

- Italian Mainstream newspapers generated from 2 to 6M interac-
tions;
- IT disinformation outlets generated no more than 500k interac-
tions each; The top-3 are a generic untrustworthy website
("sileenziefalsita.it"), the far-right website “ilprimaconziale.it”
and a law enforcement fan club ("sostenitori.it")
- only one RU website (“it.sputniknews.com”), which is technically
in Italian language but notably associated to a Russian press
agency, generated more than 100k interactions, whereas the other
- EU and US sources did not receive much attention, rarely exceed-
ing 3k interactions.
to a popular pseudo-journalistic TV program ("Le Iene") and 4 pages related to right-wing politicians, including 2 pages entitled to Matteo Salvini and 1 page entitled to Luca Zaia, governor of Veneto, one of the regions most affected by the virus. Each of these accounts generated 13 to 50 M interactions during the period of observation.

For what concerns controversial topics, we consider the Top-10 ranking according to the total number of interactions generated only by posts related to each topic (see the Appendix, Tables A.7–A.9). For what concerns 5G, we notice that most influential accounts shared only 2 to 6 related posts, but they generated 90 K to almost 2M interactions, which were accounted by 2 posts of the Italian Health Ministry. For what concerns MIGRANTS, we notice that most influential accounts generated, most of which are accounted by Matteo Salvini, leader of the right-wing Lega party. Finally, for what concerns MIGRANTS we see a larger number of posts and total interactions generated, most of which are accounted by a newspaper ("Tgcom24") and Matteo Salvini, respectively with 3.4M and 1.1M total engagement.

Similar considerations hold also when analyzing the ranking by number of posts. Overall, as shown in Fig. 1 (right), unreliable sources had a limited yet not negligible amount of engagement (1.5M) compared to news websites which convey reliable information (35.9M), in accordance with contemporary analyses [5–7]. Finally, it is interesting to notice that Disinformation sources generate relatively more engagement for the LABS topic, while the MIGRANTS topic was the most discussed among Mainstream news websites.

### 3.2. Accounts’ characteristics

For what concerns metrics for accounts involved in the analysis, such as the total engagement, total number of posts, number of members and total number of shared links, we report that in all cases their distribution approximately follows power laws, which are common in social networks dimensions [27,29] (see the Appendix). We then analyzed the total engagement generated by different accounts involved in the analysis, such as the total engagement, total number of posts, number of members and total number of shared links, we report that in all cases their distribution approximately follows power laws, which are common in social networks dimensions [27,29] (see the Appendix).

| Accounts | Posts | Accounts | Posts | Accounts | Posts | Accounts | Posts | Accounts | Posts |
|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|
| 5G       | 10937  | 25695    | 38486  | 39       | 72440  | 1588536  |
| LABS     | 5493   | 7076     | 11238  | 1958     | 15865  | 56436    |
| MIGRANTS | 5817   | 15278    | 21135  | 31       | 40175  | 715104   |
| Intersection | 3194 | 4129     | 6571   | 1232     | 9007   | 28721    |
| Union    | 5120   | 10417    | 17351  | 8        | 873432 | 873432   |
| Total    | 2299   | 2947     | 4667   | 726      | 6858   | 27715    |

We filtered out accounts with only 1 post to remove noise.

### 3.3. Polarization of accounts

To investigate the polarization [30] of accounts towards different topics we introduce a polarization score $\rho$ defined as:

$$\rho = \frac{p_c - p_n}{p_c + p_n}$$

where $p_c$ and $p_n$ are, respectively, the number of controversial and non-controversial posts of the considered account. We define a controversial post any post that contains at least one of the manually selected tokens (see the Appendix). The polarization index is constrained between $-1$, $1$.

![Fig. 2. Scatter plot of the number of members vs. number of posts (left) and number of interactions (center) for each account. Points are grouped in 100 bins to ease the visualization. (Right) Boxplot of the distribution of relative changes in the number of members/followers for all accounts, for different groups (outliers are filtered).](image)

![Fig. 3. Histogram of the polarization scores of accounts.](image)
when all the posts of an account are about controversial topics \( (p_a = 0) \) and +1, when no posts involved controversial topics \( (p_a = 0) \).

Fig. 3 shows the distribution of the polarization scores of accounts. We notice a trimodal distribution: the main peak is at \( \rho = 1 \), it represents the majority of accounts not talking about controversial topics at all; a second peak occurs at \( \rho = 0 \), it includes all accounts equally discussing controversial and non-controversial topics; the third and lower peak is at \( \rho = -1 \), it represents those accounts posting only about controversial topics.

In Fig. 4 we also show how accounts are polarized when comparing topics against each other with the same rationale, i.e., by defining \( p_a \) the number of posts about one controversial topic (e.g., 5G) and \( p_b \) the number of posts about a different controversial topic (e.g., LABS). Peaks at \( \rho = +1 \) and \( \rho = -1 \) indicate that most accounts usually do not talk about more than one controversial topic.

3.4. Linguistic analysis

In Fig. 5 we show a kernel density plot for the embedding of accounts obtained as previously described in Section 2.4. For visualization purposes, we select the first two PCA components. Red indicates controversial accounts, i.e., accounts that published at least two posts about controversial topics, whereas blue indicates the remaining ones. Note that, even if the intersection of controversial accounts is negligible (see Table 1) and the accounts are usually polarized on a single controversial topic (see Fig. 4), the embeddings of the two “classes” have similar distributions. This result suggests that controversial themes are characterized by a common lexicon, distinct from the remainder of the dataset. The aforementioned embeddings might also be suitable input feature vectors for the definition of a finer classifier able to tell apart controversial posts and accounts not relying on predefined lists of keywords and/or news sources. The definition of a similar classifier is however beyond the scope of this paper and is left to future work.

3.5. Sentiment analysis

We compute sentiments of posts using Neuraly’s “Bert-italian-cased-sentiment” model’ hosted by Huggingface [31]. It is a BERT base model [32] trained from an instance of “bert-base-italian-cased” and fine-tuned on an Italian dataset of 45 K tweets on a 3-classes sentiment analysis task (negative, neutral and positive) [33] obtaining 82% test accuracy. Previous work showed that text length can affect the classification accuracy of pre-trained models [34]. The model used in this paper, however, performs extremely well also for texts of variable length and, albeit the model was trained using short texts (i.e., tweets), it seems to benefit from the use of the entire available text (see Fig. A.5 in the Appendix). As a consequence, the sentiment analysis is obtained truncating the texts at 1960 characters – a value identified experimentally as the optimal trade-off between efficiency and accuracy, since using longer texts does not provide any measurable classification gain.

In Fig. 6 we show how the general sentiment of posts evolves during the selected months by plotting the percentage of positive and negative posts weighted by the number of shares. We remark that, even if not shown in the figure, the great part of posts is classified as neutral (81.5%). This value decreases to 78.8% when the posts are weighted by their number of shares. Positive and negative peaks can be mapped to news and events, e.g., the two main peaks of negative sentiment, occurring on January 24 and February 10, match with the first confirmed COVID-19 cases in Europe and the first confirmed 1000 deaths worldwide, while the two main peaks of positive sentiment, occurring on February 2 and March 12, correspond to the successful isolation of the virus in the “L. Spallanzani” National Institute for Infectious Diseases and the diffusion of the #andrätuttobene (“it’ll all work out”) hashtag and slogan (see Fig. A.6 in the Appendix).

7 https://huggingface.co/neuraly/bert-base-italian-cased-sentiment.
8 https://huggingface.co/dbmdz/bert-base-italian-cased.
Fig. 6. 7-days rolling average of the percentage of posts classified as positive or negative, weighted by the number of shares.

Fig. 7. Percentage of reached URLs when visiting accounts in decreasing order of degree centrality.

4. Sharing diffusion network

To better understand the patterns of sharing diffusion on Facebook, we now focus on the bipartite graph of accounts and URLs and on its projection upon the two layers, as defined in Section 2.5. Since we are particularly interested in characterizing controversial URLs and accounts, we will focus on the subgraph induced by such URLs.

4.1. Bipartite graph

By inspecting the bipartite graph we aim to investigate two related aspects of the COVID-19 infodemic on Facebook: (i) whether there are niches of accounts where possibly extreme conspiracy theories get diffused; and, conversely, (ii) whether there exists a relatively small set of accounts that, altogether, provide access to a vast majority of all available web resources.

To answer question (i), we look at the connected components of the graph. The giant component of the entire bipartite graph includes \( \approx 57\% \) of all accounts, \( \approx 88\% \) of all URLs and \( \approx 92\% \) of all edges. With a bit of simplification, this means that, limited to our dataset, more than half of all Facebook accounts draw upon a unique large pool of web content. Quite interestingly, a similar scenario emerges if we only consider the set of controversial URLs: this bipartite subgraph has 55,411 vertices (18,681 accounts and 36,730 URLs) and 81,707 edges, and its giant component includes \( \approx 72\% \) of both accounts and URLs and \( \approx 87\% \) of all edges. Components other than the giant are at least two orders of magnitude smaller in both graphs.

The isolation of specific accounts and URLs into such small components seems to emerge as a consequence of marketing strategies. For both graphs, in fact, the majority of the components consist of a small number of accounts – often, just one – sharing many different URLs (see Fig. A.7 in the Appendix). Through manual investigation, we verified that such phenomenon is oftentimes caused by a website controlling one or more Facebook pages to promote its articles. A notable case is “howtodofor.com”, which seems to use as many as 17 different one or more Facebook pages to promote its articles. A notable case that such phenomenon is oftentimes caused by a website controlling its “facebook.com” and “youtube.com” account for, respectively, 8% and 4% of all URLs and accounts.

To answer question (ii), we observe that few accounts are sufficient to reach the majority of the URLs shared in the network, as shown in Fig. 7. To reach 25% of the URLs we need 88 (0.65%) accounts, to reach 50% of the URLs we need 3,539 (26.48%) accounts, to reach 75% of the URLs we need 1,535 (11.48%) accounts, finally to reach 90% of the URLs we need 3,539 (26.48%) accounts.

4.2. Controversial URLs, domains and accounts

In this section, we consider the projection of the controversial bipartite graph upon the two layers of URLs and accounts. A naive projection leads to a URL graph with 26,705 vertices and 1,096,672 edges, and to an account graph with 13,363 vertices and 986,509 edges. This projection is useful to analyze a few macroscopic structural properties of the graph. In particular, the diameter, radius and average path length of the two graphs – 10, 5 and 3.24 for URLs, 10, 5 and 2.86 for accounts – depict a small world\(^9\); further confirmed by the global efficiency – 0.33 for URLs, 0.37 for accounts – and the clustering coefficient – 0.38 for URLs, 0.68 for accounts. This means that we observe Facebook accounts which often share common sets of URLs. At the same time, visiting a small percentage of them is enough to cover all the URLs shared in the network, as showed in Fig. 7. Altogether, this suggests that if Facebook allowed users to “jump” through groups and pages via shared URLs, they would likely get to all controversial URLs no matter the stance towards the topic. On the one hand, finding propaganda items on Facebook appears an easy task; on the other hand, debunking articles are probably equally easy to find, if the right instruments for browsing content were provided.

In line with the related literature\(^{35}\), we now focus on the statistically relevant edges of the two partitions, i.e., those edges that do not just occur as a consequence of the level of activity of groups and pages and of the “vitality” of individual news pieces. To make the computation affordable, we first lighten the bipartite graph by pruning all peripheral vertices having degree less than 10, namely, all URLs shared by less than 10 different accounts and all accounts that shared less than 10 different URLs. We then apply the validation process described in details in\(^{28}\) and summarized in Section 2.5. This process greatly reduces the size and density of the two projections, leading to a URL graph with 442 vertices and 944 edges, and an account graph with 341 vertices and 689 edges.\(^{10}\) As expected, these validated networks are also highly clustered: we computed modularity-based clusters relying on the well-known Louvain algorithm\(^{36}\), obtaining two partitions with modularity \(\approx 0.86\) for URLs and \(\approx 0.93\) for accounts. We especially aim to leverage on our classification of web domains (cf. Section 2.2) and of posts and accounts (cf. Section 2.3) to characterize these clusters. As a side result, we expect to gain insights into the possibility to infer the quality of web content solely based on where and who shared a news item.

First, in Fig. 8 we compare the ratio of Mainstream URLs with the ratio of Disinformation URLs present in each cluster, and in the entire validated URL and account graphs. We see that the ratio of deceptive URLs is very low for almost all clusters, and that in general news-related URLs are a minority. One of the reasons is the great diffusion of user generated content, made available through specific web platforms, such as Facebook itself or Youtube, that, by definition, cannot be classified as Disinformation even though they often host unreliable information. “facebook.com” and “youtube.com” account for, respectively, \(\approx 48\%\) and \(\approx 4\%\) of all URLs in our dataset.

Focusing on clusters of URLs, there are only 5 clusters having a positive, greater than the average, prevalence of Disinformation. It is especially noteworthy that one of these clusters is the giant cluster of the graph (67 URLs, 131 edges) and that it also contains several

\(^9\) A ultra-small world has \(L \propto \log \log N\), where \(L\) is the average path length and \(N\) is the number of nodes.

\(^{10}\) We used a significance level of 0.05, but we also tested 0.01 obtaining similar results.
Table 2
Top 10 URLs by PageRank.

| Label                                           | Pagerank | Disinfo categories |
|-------------------------------------------------|----------|--------------------|
| fattieavvenimenti.it/migranti-coronavirus-154... | 0.00684  | mig                |
| nextquotidiano.it/libero-esulta-per-il-coronav...| 0.00633  | 5g                 |
| youtube.com/watch?v=UT7fK4sACtw                | 0.00607  | lab                |
| facebook.com/danilottoninelli.m5s/photos/a.3947... | 0.00571  | mig                |
| gayburg.com/2020/02/lordine-de-giornalisti-aux...| 0.00565  | 5g, mig            |
| facebook.com/salviniofficial/videos/8199186318...| 0.00549  | mig                |
| quotidiano.net/cronaca/coronavirus-animali-can...| 0.00528  | lab                |
| ilgiornale.it/news/cronache/i-porti-italiani-r...| 0.00519  | mig                |
| facebook.com/nicola.morra.63/videos/5294019877...| 0.00518  | 5g                 |

Table 3
Top 10 accounts by PageRank.

| Label                                           | Pagerank | 5g% | lab% | mig% |
|-------------------------------------------------|----------|-----|------|------|
| Governo Conte - M5S - Luigi Di Maio - Di Battista  | 0.00839  | 1.97| 4.98 | 3.93 |
| Gruppo sostenitori di Vox Italia                 | 0.00686  | 3.77| 14.63| 5.54 |
| Gruppo Tutto TRAVAGLIO Forever                    | 0.00619  | 1.95| 4.71 | 5.35 |
| Amici di Diego Fusaro                             | 0.00602  | 4.05| 10.81| 4.05 |
| MATTEO SALVINI, E GIORGIA MELONI PER UNTITALIA    | 0.00555  | 1.02| 4.1  | 19.28|
| DOTT. ANTONIETTA MORENA GATTI E DOTT. STEFANO MONTANARI PATRIMONIO DELL’UMA | 0.00546 | 7.87| 15.75| 1.57 |
| Vaccini Politi. Rimozione dal commercio dei prodotti vaccinali contaminati | 0.00532 | 3.86| 16.02| 3.37 |
| Fan di Vox Italia - Diego Fusaro                  | 0.00496  | 5.1 | 10.97| 8.93 |
| Parlaimone con... Paolo Barnard                   | 0.00496  | 8.33| 19.44| 8.33 |
| Salvini premier la rivoluzione del buonsenso idee cuore e coraggio | 0.00496 | 1.6 | 7.19 | 14.78 |

Fig. 8. Prevalence of Mainstream and Disinformation domains in different clusters of the validated URL graph and account graph. The marker size is proportional to the cluster size.

Mainstream URLs, most of which belong to either “ilgiornale.it” or “liberoquotidiano.it”, two outlets notably close to the Italian sovereign right parties. This hints to the fact that successful propaganda builds on the support of well-known and accredited media. The two clusters of URLs with the greatest ratio of Disinformation are instead very small clusters (6 URLs) with a single known Disinformation URL. However, the other 5 URLs of one of these two clusters all belong to the same domain, “oasisana.com”, which, after manual inspection, resulted being a “free and natural information” website, mostly focused on anti-5G propaganda. This finding speaks in favor of the possibility to use source-based labeling and network-based clustering combined to identify other previously unknown Disinformation domains.

For what regards accounts, the clusters with a significant share of Disinformation URLs are composed of a few groups and pages that cooperate to diffuse news pieces produced by a well-defined and small set of domains. Especially remarkable are:

- A set of 6 accounts with nationalism-referring names that share content from “tg24-ore.com”, “curiosity-online.com” and “howtodofor.com”.
- Two pages close to the “Lega Nord” (“Lega - Salvini Premier News” and “Notizie Lega Nord”), mostly brought together by the diffusion of hundreds of news pieces from “ilfattoquotidiano.it” and “ilgiornale.it”. Quite interestingly, the most diffused domain by the page “Notizie Lega Nord” is “it.sputniknews.com”, with almost 2 K URLs shared, which is the Italian version of the well-known Russian propaganda agency Sputnik.
- A cluster of 8 anti-5G groups and pages linked together by URLs from the aforementioned “oasisana.com” and all contributing to the diffusion of conspiracy theories through videos shared on Youtube or directly on Facebook.
- A cluster of 15 fan groups of sovereign right leaders Matteo Salvini and Giorgia Meloni that, within a mix of news from Disinformation and Mainstream media, re-share video produced by the two leaders themselves on Facebook.
These findings suggest that propaganda upon controversial topics is driven by well-defined ideological affiliation, and that a few political parties, their leaders and their supporting Mainstream media have a precise role in shaping the public opinion. It also shows, however, that anti-scientific propaganda is apparently cross-ideological.

To further characterize the obtained clusters, in Fig. 9 we show the prevalence of each of the three controversial topics in different clusters of the validated URL graph and account graph. The line width is proportional to the cluster size.

Fig. 9. Prevalence of controversial topics in different clusters of the validated URL graph and account graph. The line width is proportional to the cluster size.

5. Conclusions

In this paper we investigated online conversations about COVID-19 and related controversial topics on Facebook, during a period of 4 months and analyzing more than 1.5M posts shared by almost 80k groups and pages. We first noticed that discussions on controversial topics, which had a smaller volume of interactions compared to the pandemic in general, induced polarized clusters of accounts in terms of both topic coverage and lexicon. We then observed that, in accordance with recent literature, sources of supposedly reliable information had a higher engagement compared to websites sharing unreliable content. However, we also realized the limitations of source-based approaches when analyzing an information ecosystem wherein user generated content has a paramount role. Further, we highlighted a "small-world effect" in the sharing network of URLs, with the result that users on Facebook who navigate on a limited set of pages/groups can be potentially exposed to a wide range of content, from extreme propaganda to verified information. Finally, we only considered statistically validated links between URLs and accounts to discover a significant level of coordination for the diffusion of propaganda and disinformation. In this network, the central role is taken by popular groups, in contrast with popular pages being those generating the greatest engagement.

Future directions of research include further investigating the differences in the activity of groups and pages which focus on controversial topics. In particular, we aim to understand whether language differences might be effectively employed to distinguish accounts who were particularly (in)active on specific subjects, and to extend the analysis of reliable vs. unreliable information to platforms for video and image sharing such as YouTube and Instagram.

CRediT authorship contribution statement

Stefano Guarino: Data curation, Conceptualization, Methodology, Software, Visualization, Investigation, Validation, Writing - original draft, Writing - review & editing. Francesco Pierri: Conceptualization, Methodology, Software, Visualization, Investigation, Validation, Writing - original draft, Writing - review & editing. Marco Di Giovanni: Conceptualization, Methodology, Software, Visualization, Investigation, Validation, Writing - original draft, Writing - review & editing. Alessandro Celestini: Conceptualization, Methodology, Software, Visualization, Investigation, Validation, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Table A.1
List of Italian Mainstream news sources.

| Source                  |
|-------------------------|
| repubblica.it           |
| tгcom24.mediaset.it     |
| ilfattoquotidiano.it    |
| ansa.it                 |
| lastampa.it             |
| huffingtonpost.it       |
| leggo.it                |
| nanopress.it            |
| ilgazzettino.it         |
| ilpost.it               |
| letters43.it            |
| iltempo.it              |
| avvenire.it             |
| corriere.it             |
| ilmessaggero.it         |
| fanpage.it              |
| ilgiornale.it           |
| ilsole24ore.com         |
| quotidiano.net          |
| ilmattino.it            |
| tg24.sky.it             |
| rainews.it              |
| agi.it                  |
| adnkronos.com           |
| today.it                |

Table A.2
List of Italian Disinformation news sources.

| Source                  |
|-------------------------|
| agenpress.it            |
| essere-informati.it     |
| voxnews.info            |
| direttraneu24.com       |
| terrarealtetime.blogspot.com |
| pandoravt.it            |
| notiziariomacapitale.blogspot.com |
| webt24.com              |
| hackthematrix.it        |
| tg24-ore.com            |
| corrieredeicorsaro.it   |
| liberedee.org           |
| ilfattoquotidiano.it    |
| compressamento.blogspot.com |
| nibiru2012.it           |
| lonesto.it              |
| ilprimotonazionale.it   |
| comedonchisciotti.org  |
| ununiverso.it           |
| ilviostroppensiero.it   |
| sky tg24news.it         |
| altrarealta.blogspot.com|
| adesosbasta.org         |
| tuttiscrinidegliimmigrati.com |
| tankerenemy.com         |
| freenondarrevolution.wordpress.com |
| catenaumana.it          |
| laveritadinconacoro.altervista.org |
| ecplanet.org            |
| ilsaperepotere2.blogspot.com |
| lettoquotidiano.it      |
| sostentitori.info       |
| italianoosveglia.com    |
| madterra.myblog.it      |
| ilpunsutodismistero.it  |
| disinformazione.it      |
| interagisco.it          |
| conoscezealconfine.it   |
| neovitruvian.wordpress.com |
| accademiadellaliberta.blogspot.com |
| byoblu.com              |
| mag24.es                |
| zapping2017.myblog.it   |
| 5stel lenews.com        |
| tmcrew.org              |
| planetax.wordpress.com  |
| jedanews.it             |
| skynew.it               |

Table A.3
List of US Disinformation news sources.

| Source                  |
|-------------------------|
| drudgereport.com        |
| theblaze.com            |
| pakalertpress.com       |
| geoeengineeringwatch.org|
| infowars.com            |
| collective-evolution.com|
| disclose.tv             |
| govtslaves.info         |
| thefreethoughtproject.com|
| breitbart.com           |
| worldtruth.tv           |
| aktivismpost.com        |
| worldnewsdailyreport.com|
| naturalnews.com         |
| 21stcenturywire.com     |
| prisonplanet.com        |
| beforeitsnews.com       |
| thedailysheeple.com     |
| thedailybeeply.com      |
| realfarmacy.com         |

Appendix

The complete list of Mainstream and Disinformation news sources used in this paper is reported in Tables A.1–A.5. The Top-10 ranking of such news websites by total engagement is reported in Fig. A.2.
Fig. A.2. Top-10 ranking of news sources per different news domain according to the total engagement generated. In clockwise order from top left we show US, RU, IT, EU disinformation sources and finally IT Mainstream sources.

Fig. A.3. Complementary cumulative distribution function for several metrics. We show all accounts and according to different topics.
The lists of Italian words used to define the three controversial topics are the following (the queries included also the feminine and plural forms, omitted here for the sake of clarity):

- “5G”: elettromagnetismo, onda, radiazione, wireless;
- “LABS”: cavia, espetimento, sperimentato, sperimentazione;
- “MIGRANTS”: africa, barcone, clandestino, extracomunitario, immigrato, islam, musulmano, negro, niger, ONG, profugo, senegal, straniero

In Fig. A.1 we show the time series of the daily engagement, in total and for each controversial topic, whereas in Fig. A.3 we plot the complementary cumulative distribution of total engagement (No. reactions), total number of posts (No. posts), total number of shared links (No. links) and number of members (No. members), for all accounts and for the accounts involved in each controversial topic.

The distribution of the relative change in the number of members/followers of the accounts during the observation period is depicted as a boxplot in Fig. A.4. The Top-10 ranking of the accounts by total engagement, considering either all accounts or the accounts involved in a specific controversial topic, is reported in Tables A.6–A.9.

In Fig. A.5 we show the dependence upon the text truncation length of the average accuracy and $F_1$ score of the sentiment classification model used in this paper. The scores increase as we increase the truncation length, even if the resulting sentences are longer than the maximum length of sentences from our training set (280 characters).

To perform this analysis we used a Tripadvisor dataset of 28754 Italian reviews of hotels and restaurants, with an average length of about 700 characters. It can be found and downloaded at the following link (development dataset): http://dbdmg.polito.it/wordpress/wp-content/uploads/2020/01/dataset_winter_2020.zip.

The time series of the daily number of posts including the #andràtuttobene (“it’ll all work out”) hashtag is shown in Fig. A.6. We notice a clear peak on March 12, which is reflected in a rise of positive sentiment in the dataset.

In Fig. A.7 we compare the number of URLs and the number of accounts belonging to each connected component of the bipartite

---

**Table A.6**

| Account                  | No. posts | No. interactions |
|--------------------------|-----------|------------------|
| Le Iene                  | 256       | 4986/9767        |
| Fanpage.it               | 3043      | 4923/4295        |
| Corriere della Sera      | 2119      | 2800/4758        |
| Vittorio Sgarbi          | 98        | 2186/5677        |
| Tgcom24                  | 2559      | 1965/4257        |
| Sky TG24                 | 1494      | 1925/9609        |
| Notizie.it               | 2513      | 1690/7208        |
| Matteo Salvini           | 215       | 1671/9496        |
| Luca Zaia                | 207       | 1606/8773        |
| Lega - Salvini Premier   | 1386      | 1318/7120        |

**Table A.7**

| Account                  | No. posts | No. interactions |
|--------------------------|-----------|------------------|
| Ministero della Salute    | 2         | 1863/319         |
| Nicola Morra              | 2         | 982/903          |
| Che tempo che fa          | 4         | 383/441          |
| Quarto Grado              | 2         | 294/430          |
| Sfera                     | 3         | 1216/35          |
| Lorenzo Tosa              | 2         | 1085/86          |
| Abolizione del suffragio universale | 3 | 1084/91 |
| Il Sole 24 ORE            | 6         | 1044/33          |
| Angelo DURÓ               | 2         | 928/55           |
| Fondazione Poliambulanza Istituto | 2 | 904/81 |

**Table A.8**

| Account                  | No. posts | No. interactions |
|--------------------------|-----------|------------------|
| Matteo Salvini           | 19        | 1443/184         |
| Nicola Porro             | 12        | 456/688          |
| Silvia Sardone           | 9         | 416/352          |
| Lega - Salvini Premier   | 55        | 382/654          |
| Medici con l'Africa Cuamm| 20        | 331/577          |
| Tg3                      | 18        | 280/473          |
| Sky TG24                 | 7         | 220/786          |
| Local Team               | 3         | 200/64           |
| Abolizione del suffragio universale | 7 | 173/648 |
| Fanpage.it               | 26        | 169/921          |

**Table A.9**

| Account                  | No. posts | No. interactions |
|--------------------------|-----------|------------------|
| Tgcom24                  | 58        | 340/8934         |
| Matteo Salvini           | 12        | 115/7440         |
| Kiko.Co                  | 2         | 826/064          |
| Luca Zaia                | 9         | 603/91           |
| Gianni Simioni           | 3         | 577/797          |
| Vincenzo De Luca         | 25        | 535/635          |
| Tg1                      | 3         | 489/346          |
| Sky TG24                 | 20        | 442/357          |
| Il Messaggero.it         | 21        | 406/447          |
| Tg3                      | 16        | 382/525          |

**Fig. A.4** Boxplot of the distribution of relative changes in the number of members/followers for all accounts, groups and pages.

**Fig. A.5** Average accuracy and $F_1$ score of the sentiment classification model when texts are truncated at different lengths.

**Fig. A.6** Number of posts with the slogan “andràtuttobene” (“it’ll all work out”).
graph of accounts and URLs. We consider both the entire graph and the subgraph induced by controversial topics.

References

[1] DPCM, Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell’emergenza epidemiologica da COVID-19, Gazzetta Ufficiale 62 (09-03-2020) (2020).

[2] G. Bonacors, F. Pierri, M. Cinelli, F. Porcelli, A. Galeazzi, A. Flori, A.L. Schmidt, C.M. Valensise, A. Scala, W. Quattrociocchi, F. Pammolli, Economic and social consequences of human mobility restrictions under COVID-19, Proc. Natl. Acad. Sci. (2020).

[3] A. Galeazzi, M. Cinelli, G. Bonacors, F. Pierri, A.L. Schmidt, A. Scala, F. Pammolli, W. Quattrociocchi, Human mobility in response to COVID-19 in France, Italy and UK, 2020, arXiv:2005.06341.

[4] J. Jarocostas, How to fight an infodemic, Lancet 395 (10225) (2020) 676.

[5] R. Gallotti, F. Valle, N. Castaldo, P. Sacco, M. De Domenico, Assessing the risks of “infodemics” in response to COVID-19 epidemics, 2020, pp. 1–29, arXiv:2004.03997.

[6] K.-C. Yang, C. Torres-Lugo, F. Menczer, Prevalence of low-credibility information on Twitter during the COVID-19 outbreak, 2020, arXiv:2004.14484.

[7] 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, 2020, pp. 1–18, arXiv:2003.05004.

[8] H. Allocit, M. Gentzow, Social media and fake news in the 2016 election, J. Econ. Perspect. 31 (2) (2017) 211–236.

[9] D.M.J. Lazer, M.A. Baum, Y. Benkler, A.J. Berinsky, K.M. Greenhill, F. Menczer, M.J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schud- son, S.A. Somani, C.R. Sunstein, E.A. Thorton, D.J. Watts, J.L. Zittrain, The science of fake news, Science 359 (6368) (2018) 1094–1096, http://dx.doi. org/10.1126/science.aao2998, arXiv:http://science.sciencemag.org/content/359/6368/1094.full.pdf, Ut:http://science.sciencemag.org/content/359/6368/1094.

[10] C. Shao, G.L. Campiglia, O. Varol, K.-C. Yang, A. Flammini, F. Menczer, The spread of low-credibility content by social bots, Nature Commun. 9 (1) (2018) 4787.

[11] F. Pierri, S. Ceri, False news on social media: a data-driven survey, ACM Sigmod Rec. 48 (2) (2019).

[12] R.K. Nielsen, N. Newman, R. Fletcher, A. Kalogeropoulos, Reuters Institute Digital News Report 2019, Report of the Reuters Institute for the Study of Journalism, 2019.

[13] G. Caldarelli, R. De Nicola, F. Del Vigna, M. Petrocchi, F. Saracco, The role of bot squads in the political propaganda on Twitter, 2019, arXiv preprint arXiv:1905.12667.

[14] S. Guarnino, N. Trino, A. Chessa, G. Riotta, Beyond fact-checking: Network analysis tools for monitoring disinformation in social media, in: H. Cherifi, S. Gaito, J.F. Mendes, E. Moro, L.M. Rocha (Eds.), Complex Networks and their Applications VIII, Springer International Publishing, Cham, 2020, pp. 436–447.

[15] S. Guarino, N. Trino, A. Celestini, A. Chessa, G. Riotta, Characterizing networks of propaganda on Twitter: a case study, 2020, arXiv preprint arXiv:2005.10004.

[16] M. Del Vicario, S. Gaito, W. Quattrociocchi, M. Zignani, 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 (DSSA), IEEE, 2017, pp. 648–657.

[17] F. Pierri, A. Artoni, S. Ceri, Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections, PLoS One 15 (1) (2020) e0227821.

[18] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 U.S. presidential election, Science 363 (6425) (2019) 374–378, http://dx.doi.org/10.1126/science.aau2706, arXiv:http://science.sciencemag.org/content/363/6425/374.full.pdf, Ut:http://science.sciencemag.org/content/363/6425/374.

[19] A. Bovet, H.A. Makse, Influence of fake news in Twitter during the 2016 US presidential election, Nature Commun. 10 (1) (2019) 7, http://dx.doi.org/10. 1038/s41467-018-07761-2.

[20] F. Pierri, C. Piccardi, S. Ceri, Topology comparison of Twitter diffusion networks effectively reveals misleading news, Sci. Rep. 10 (2020) 1372.

[21] F. Pierri, The diffusion of mainstream and disinformation news on Twitter: the case of Italy and France, in: Companion Proceedings of the Web Conference 2020 (WWW ’20 Companion), 2020.

[22] F. Giglietto, L. Iannelli, L. Rossi, A. Valeriani, N. Righetti, F. Carabini, G. Marino, S. Usai, E. Zurovac, Mapping Italian news media political coverage in the lead-up to 2018 general election, 2018, Available at SSRN: https://ssrn.com/abstract=3030030.

[23] CrowdTangle Team, Crowdtriangle. Menlo park, CA: Facebook, 2020, Available at: https://www.crowdtangle.com.

[24] M. Honnibal, I. Montani, Spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing, 2017, http://dx.doi.org/10.5281/zenodo.1212304, in press.

[25] S. Bird, E. Klein, E. Loper, Natural Language Processing with Python, first ed., O’Reilly Media, Inc., 2009.

[26] A.-L. Barabási, et al., Network Science, Cambridge university press, 2016.

[27] M. Newman, Networks, Oxford university press, 2018.

[28] F. Saracco, M.J. Straka, R. Di Clemente, A. Gabrielli, G. Caldarelli, T. Squarzini, Inferring multipartite projections of bipartite networks: an entropy-based approach, New J. Phys. 19 (5) (2017) 053022.

[29] A.-L. Barabási, Network Science, Cambridge university press, 2016.

[30] M. Cinelli, G.D.F. Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, Echo chambers on social media: A comparative analysis, 2020, arXiv preprint arXiv: 2004.09603.

[31] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Fountowicz, J. Brew, HuggingFace’s Transformers: State-of-the-art natural language processing, 2019, arXivXiv:1910.03771.

[32] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv:1810.04805.

[33] B. Pasile, F. Cutugno, M. Nissim, V. Patti, R. Sprugnoli, EVALITA 2016: Overview of the 5th evaluation campaign of natural language processing and speech tools for Italian, in: CEUR Workshop Proceedings, vol. 1749, 2016.

[34] R.K. Amplayo, S. Lim, S.-w. Hwang, Text length adaptation in sentiment classification, in: W.S. Lee, T. Suzuki (Eds.), Proceedings of the Eleventh Asian Conference on Machine Learning, in: Proceedings of Machine Learning Research, 101, PMLR, Nagoya, Japan, 2019, pp. 646–661, http://proceedings.mlr.press/p101/amplayo19a.html.

[35] M. Cinelli, G. Caldarelli, A. Galeazzi, M. Petrocchi, F. Saracco, The role of propaganda on Twitter: a case study, 2020, arXiv preprint arXiv:2005.10004.

[36] G. Caldarelli, R. De Nicola, F. Del Vigna, M. Petrocchi, F. Saracco, The role of bot squads in the political propaganda on Twitter, 2019, arXiv preprint arXiv:1905.12667.

[37] C. Becatti, G. Caldarelli, R. Lambiote, F. Saracco, Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections, Palgrave Commun. 5 (1) (2019) 1–16.

[38] V.D. Blondel, J.-L. Guillaume, R. Lambiote, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008.

[39] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: bringing Order to the Web, Tech. rep., Stanford InfoLab, 1999.