Applied Network Science

RESEARCH Open Access

Characterizing networks of propaganda on twitter: a case study

Stefano Guarino[1*], Noemi Trino[2], Alessandro Celestini[1], Alessandro Chessa[2,3] and Gianni Riotta[2]

*Correspondence:
s.guarino@iac.cnr.it
[1]Institute for Applied Mathematics, National Research Council, Rome, Italy
Full list of author information is available at the end of the article

Abstract

The daily exposure of social media users to propaganda and disinformation campaigns has reinvigorated the need to investigate the local and global patterns of diffusion of different (mis)information content on social media. Echo chambers and influencers are often deemed responsible of both the polarization of users in online social networks and the success of propaganda and disinformation campaigns. This article adopts a data-driven approach to investigate the structuration of communities and propaganda networks on Twitter in order to assess the correctness of these imputations. In particular, the work aims at characterizing networks of propaganda extracted from a Twitter dataset by combining the information gained by three different classification approaches, focused respectively on (i) using Tweets content to infer the "polarization" of users around a specific topic, (ii) identifying users having an active role in the diffusion of different propaganda and disinformation items, and (iii) analyzing social ties to identify topological clusters and users playing a "central" role in the network. The work identifies highly partisan community structures along political alignments; furthermore, centrality metrics proved to be very informative to detect the most active users in the network and to distinguish users playing different roles; finally, polarization and clustering structure of the retweet graphs provided useful insights about relevant properties of users exposure, interactions, and participation to different propaganda items.

Keywords: Propaganda networks, Polarization, Centrality, Clustering

Introduction

The 2016 US presidential election veritably marked the transition from an age of 'post-trust' (Löfstedt 2005), to an era of 'post-truth' (Higgins 2016), with contemporary advanced democracies experiencing a rise of anti-scientific thinking and reactionary obscurantism, ranging from online conspiracy theories to the much-discussed "death of expertise" (Nichols 2017). The long-standing debate about the relationship between media and public good has been reinvigorated: the initial euphoria about the "openness" of the Internet (Lévy 2002) has been taken over by a widespread concern that social media may instead be undermining the quality of democracy (Tucker et al. 2018). Media outlets, public officials and activists are supplying citizens with different, often contradictory "alternative facts" (Allcott and Gentzkow 2017). In this context, social media platforms

would be fostering "selective exposure to information", with widespread diffusion of "echo chambers" and "filter bubbles" (Sunstein 2001; Pariser 2011). Propaganda actions may be now more effective than ever, representing a major global risk, possibly able to influence public opinion enough to alter election outcomes (Van der Linden et al. 2017; Shao et al. 2018; Guess et al. 2019).

As a first step towards the disruption of these networks of propaganda, researchers have been trying to model the social mechanisms that make users fall prey of partisan and low-quality information. From a psychological point of view, news consumption is mainly governed by so-called "informational influence", "social credibility", "confirmation bias" and "heuristic frequency" (Shu et al. 2017; Del Vicario et al. 2017). This means that social media users tend to shape their attitude, belief or behavior based on arguments provided in online group discussions, using popularity as a measure of credibility, privileging information that confirms their own prior beliefs and/or that they hear regularly. These phenomena are exacerbated by the general incapability of making good use of the great amount of available information, a problem which can be modeled relying on the dualism of information overload *vs.* limited attention (Qiu et al. 2017), or on the principles of information theory and (adversarial) noise decoding (Brody and Meier 2018). However, there is still a lack of evidence in the literature regarding the processes that lead to the structuration of digital ecosystems where polarized and unverified claims are especially likely to propagate virally. Are these a natural consequence of the existence of communities with homogeneous beliefs – i.e., echo chambers – and of the organized actions of "propaganda agents", or are we missing a piece?

To provide a first answer to this and other related questions, the present paper takes a data-driven approach. Specifically, we aim at demonstrating the importance of characterizing networks of propaganda on Twitter by combining the information gained by three different classification approaches: (i) using the content of tweets to determine users' "polarization" with respect to a main theme of interest; (ii) telling apart users having an active role in the diffusion of different propaganda and disinformation items related to that theme; (iii) analyzing social ties to identify topological clusters and users playing a "central" role in the network. Our main goal is addressing the following research questions:

- Is modularity-based network clustering "stable" or are the patterns of cohesion among users dependent of the topics of discussion? In other terms, is the exposure/participation to propaganda of a given user a direct consequence of his/her own global interactions with other users?
- Can we use centrality metrics for detecting users playing specific roles in the production-diffusion chain of propaganda? If yes, what metrics should we mostly rely on? And are these users "consistently" involved in the diffusion of related yet different propaganda items?
- What is the role of polarization in the analysis? How shall we use the available information about the political/social "goal" of a propaganda item to enrich the graph-based analysis of the corresponding network of propaganda?

Our methodology will be applied to a case study concerning the constitutional referendum held on December 4, 2016 in Italy, by means of a dataset composed of over 1.3 millions tweets. As a side result, we will provide insights regarding the reasons of the

success of specific propaganda items and the existence of "propaganda hubs" and "authorities", i.e., accounts that are critical in fostering propaganda and spreading disinformation campaigns.

Related work

As reported by a recent Science Policy Forum article (Lazer et al. 2018), stemming the viral diffusion of fake news largely remains an open problem. The body of research work on fake news detection is vast and heterogeneous: linguistics-based techniques (Markowitz and Hancock 2014; Feng et al. 2012; Feng and Hirst 2013) coexist with network-based techniques (Ciampaglia et al. 2015; Papacharissi and de Fatima Oliveira 2012; Karadzhov et al. 2017) as well as machine-learning-based approaches (Castillo et al. 2011; Zubiaga et al. 2018). Yet, (semi-)automatic debunking seems not an adequate response if considered alone (Margolin et al. 2018; Shin and Thorson 2017). Experimental evidence confirms the general perception that, on average, fake news get diffused farther, faster, deeper and more broadly than true news (Silverman and Singer-Vine 2016). Users are more likely to share false and polarized information and to share it rapidly, especially when related to politics (Vosoughi et al. 2018), while the sharing of fact-checking content typically lags that of fake news by at least 10 h (Shao et al. 2016). Furthermore, debunking is often associated to counter-propaganda and disseminated online through politically-oriented outlets, thus reinforcing selective exposure and reducing consumption of counter-attitudinal fact-checks (Shin and Thorson 2017). Besides the technical setbacks, the existence of the so-called "continued influence effect of misinformation" is widely acknowledged among socio-political scholars (Skurnik et al. 2005), thus questioning the intrinsic potential of debunking in contrasting the proliferation of fake news.

In this regard, the efforts deployed by major social media platforms seem insufficient. As of 2017, Twitter – the most widely studied of such platforms – expressed an alarmingly shallow stance towards disinformation, stating that bots are a "positive and vital tool" and that Twitter is by nature "a powerful antidote to the spreading of false information" where "journalists, experts and engaged citizens can correct and challenge public discourse in seconds" (Crowell 2017). In the meanwhile, based on two millions retweets produced by hundreds thousands accounts in the six months preceding the 2016 US presidential election, researchers were coming to the conclusion that the core of Twitter's interaction network was nearly fact-checking-free while densely populated of social bots and fake news (Shao et al. 2018).

Characterizing misinformation and propaganda networks on social media thus recently emerged as a primary research trend (Subrahmanian et al. 2016; Shao et al. 2018; Bovet and Makse 2019). Data collected on social media are paramount for understanding disinformation disorders (Bovet and Makse 2019): they are instrumental to analyze the global and local patterns of diffusion of unreliable news stories (Allcott and Gentzkow 2017) and, to a broader level, to understand the relevance of propaganda on public opinion, possibly incorporating thematic, polarity or sentiment classification (Vosoughi et al. 2018), thus unveiling the structure of social ties and their impact on (dis)information flows (Bessi and Ferrara 2016). Investigating the relation between polarization and information spreading has also been shown to be instrumental for both uncovering the role of disinformation in a country's political life (Bovet and Makse 2019) and predicting potential targets for hoaxes

and fake news (Vicario et al. 2019). Finally, recent work used network-based features as instruments to describe, classify and compare the diffusion networks of different disinformation stories as opposed to "main-stream" news, making a promising step towards text-independent fake news detection (Pierri et al. 2020).

A relevant issue emerging from the literature is quantifying the representativeness of data extracted from real-time social media in general, and more specifically from Twitter, when these data are used to forecast opinion trends and vote shares in elections. In particular, the socio-demographic composition of Twitter users may be not representative of the overall population and may thus manifest different political-preferences from non-Twitter users (Bakker and De Vreese 2011; Burckhardt et al. 2016). This potential mismatch could be accompanied by a self-selection bias: as some scholars showed (Ceron et al. 2016), the largest number of comments is often produced by the more active and politically mobilitated users, while the vast majority of accounts has a limited activity (Gayo-Avello et al. 2011). Nonetheless, the main goal of this paper is making one step forward in the understanding of the role of propaganda in shaping the political debate in Italy. To this end, Twitter is extremely representative: it is in fact the reference social media in Italy to discuss political issues. Investigating to which extent our findings may be extended to the Italian population at large is left to future work.

Background

After the crucial 2013 election, that had imposed an unprecedented tri-polar equilibrium in the Italian political scenario, the 2016 referendum determined the collapse of the entire political scene, with the defeat of the center-left "Democratic Party" and the successive resignation of its leader and head of government, Matteo Renzi, architect of the consultation. The government reform was in fact strongly defeated, with "NO" percentages at 59.12% and "YES" at 30.88%. Offline trends showed how political polarisation and divisions among party leaders fostered the grassroots activism of the YES and NO front committees, reinforcing opposite views regarding the reform. The NO faction was a composite formation supported by both left-wing and right-wing parties, with alternative yet sometimes overlapping political justifications. Subsequently, the 2018 elections sanctioned the major rise of two euro-skeptic and populist formations, "5 Stars Movement" and "The Northern League", who were the main actors of opposition to the 2016 referendum.

The constitutional referendum offered to these rising parties an extraordinary window of opportunity in propaganda building, by imposing carefully selected instrumental news-frames and narratives and using social media as strategic resources for community-building and alternative agenda setting. Social media – and Twitter in particular – have in fact constituted a strategic tool for newly born political parties, that through the activation of the two-way street mediatization could incorporate their proposals into conventional media, still maintaining a critical, even conspiratorial attitude towards traditional media (Alonso-Muñoz and Casero-Ripollés 2018; Schroeder 2018). More generally, the dichotomous structuration of referendum offered to both political alignments the chance to align the various issues along a pro/anti-status-quo spectrum. The cleavage was strategically used by both coalitions, which adopted opposite frames to stress their position:

- on the one hand, the referendum was framed as a tool of "rottamazione", the process of political renovation at the center of Renzi's political agenda;
- on the other one, on the NO front, it was inserted in the broader cleavage between anti-parties and traditional parties, pointed as an expression of old interests and privileges.

Data collection

For data collection we relied on Twitter's Streaming API, scraping tweets containing any combination of the following hashtags: "#ReferendumCostituzionale", "#IoVotoNO", "#SIcambia", "#SIRiforma", "#Italiachedicesì", "#Italiachedicesi", "#bastaunsi", "#referendum", "#costituzione", "#riformacostituzionale", "#famiglieperilno", "#bastaunsì", "#bastaunsi", "#referendumsociali". These are a mix of "trending" hashtags, official hashtags of the referendum campaign, and popular hashtags used by the supporters of the two fronts. Data was collected for the six months preceding the referendum, that is, from July 05, 2016, to December 04, 2016, but we only consider the tweets dated from November 01 in this paper in order to focus on the most relevant part of the campaign.

Propaganda items

Following the literature, in order to identify the main topics and themes of disinformation of the political campaigning we relied on the activity of fact-checking and news agencies who reported lists of (dis)information news stories that went viral during the referendum campaign. Mostly based on the work by fact-checking web portal *Bufale.net* (Mastinu 2016), online newspaper *Il Post* (Post 2016), and political fact-checking agency *Pagella Politica*[1] (Politica 2016), we were able to identify twelve main stories, including both general theories and very specific news pieces. To widen the scope of the analysis, we considered news, theories and topics of discussion that could be associated to information disorders in its broader sense. This includes *factual* (i.e., verifiably true/false) claims as well as stories (e.g., hearsays, rumors and conspiracy theories) that cannot be deemed true/false with certainty, with no distinction between deliberate and organized disinformation/propaganda and unintentionally propagated misinformation.

Differently from related work (Pierri et al. 2020) that used the presence of a specific url for collecting tweets associated to a news story of interest, we set up a custom query in order to search our dataset for tweets that discuss a given topic in a broader sense. For each of the twelve propaganda items considered, we manually selected relevant textual content related to that story – news pieces, tweets, work of debunking agencies – from which we extracted a suitable keyword-based query. An example of such queries is the following (corresponding to what will be later denoted PI2):

('illegittimo' OR 'illeggittimo' OR 'illegal' OR 'non eletto') AND ('parlamento' OR 'governo' OR 'renzi' OR 'presidente')

The query is enriched with synonyms – as in *('illegittimo' OR 'illeggittimo' OR 'illegal' OR 'non eletto')* – that take into account singular/plural forms, different jargon, and, possibly, frequent spelling errors. With the terminology of information retrieval, these synonyms are expected to increase the recall of our filters. On the other hand, to assess the precision of the filters we manually verified a sample of 200 tweets per filter, finding that all of them

[1] Pagella Politica is partner of the EU H2020 SOMA Project.

where somewhat related to the corresponding propaganda item. The size of this sample, albeit limited, must be commensurate with the total number of tweets matching each filter, which is in the order of a few thousands. It is worth noticing that we do not aim at perfect accuracy; rather, as any query-based filter, the goal was collecting a sufficiently large and significant sample of tweets for each propaganda item.

In a previous work we classified these stories into four categories (Guarino et al. 2019), by distinguishing entirely fabricated content from manipulated items and broader propaganda pieces. Here, we decided to focus upon the four most shared Propaganda Items (PI), and namely:

PI1 A newspiece about alleged vote rigging organized by government forces;
PI2 A second item framing the referendum as the political product of an illegitimately elected parliament and/or government;
PI3 A third news, claiming that victory of the YES would make Italy yield national sovereignty to EU institutions (especially referring to an hidden clause in art.117);
PI4 A fourth - more general - piece supporting the claim that a victory of the YES would have caused a shift towards authoritarianism.

All the most diffused news items can be broadly located along the spectrum of different arguments of conspiracy theories, traditionally driven by a belief that a powerful group of people is manipulating the public, while concealing their activities. As some scholars have demonstrated (Castanho Silva et al. 2017), conspirationism is associated with different sub-dimensions of populist attitudes-people-centrism, anti-elitism, and a good-versus-evil view of politics-, with coup d'état attempts and secret plots organized by political élites to gain further power or consolidate their privilege or the explicitly plot to notch the integrity of the electoral process by gaining unauthorized access to voting machines and altering voting results.

Classification of tweets and users

After having identified the most relevant news-pieces in our dataset, we aimed at gaining a better understanding of users in our dataset and the relation between polarization and disinformation. To classify the stance of each tweet with respect to the referendum question, we adopted a semi-automatic self-training process, described more in detail in (Guarino et al. 2019). The underlying idea is that political exchanges in social-media platforms exhibiting "a highly partisan community structure" with "homogeneous clusters of users who tend to share the same political identity" (Conover et al. 2011). This is reflected on Twitter by the usage of different patterns of hashtags by supporters of opposite factions (Becatti et al. 2019). We therefore built a hashtag graph, selecting the top 30 hashtags by weighted degree (i.e., with the greatest number of co-occurrences with other hashtags). Among them, we identified a set of generic and/or out-of-context hashtags that could have been detrimental to identifying clear and meaningful clusters, namely: "#referendum", "#referendumcostituzionale", "#photo", "#riformacostituzionale", "#costituzione", "#4dicembre", "#trendingtopic" and "#1w1l". Pruning these hashtags indeed increased the modularity of the clustering. The rationale was to mimic the removal of stopwords or very frequent words in order to improve the quality of topic modeling. Louvain's algorithm was then applied to cluster such hashtags based on their mutual co-occurrence patterns. We found the two greatest clusters to clearly identify the YES and NO fronts, thus we

used hashtags in these clusters to extract a training set composed of tweets labelled as follows: -1 (NO) if the tweet only contains hashtags from the NO cluster; $+1$ (YES) if the tweet only contains hashtags from the YES cluster; 0 (UNK) if the tweet contains a mix of hashtags from the two clusters.

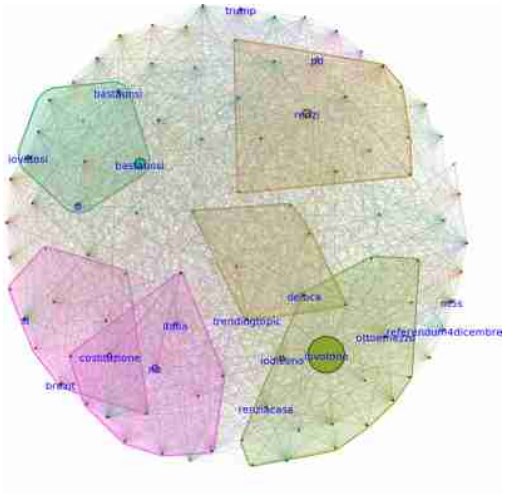
To extend the labeling to all tweets in the dataset, we defined a text-based classifier. The classifier may be tuned to represent tweets using tf-idf vectors, doc2vec (Le and Mikolov 2014), or a combination of both, and to use either Logistic Regression or a Gradient Boosting Classifier. We tested any possible combination and selected the overall best performing one, namely, a Gradient Boosting Classifier using doc2vec feature vectors. As classification score we used the mean accuracy on 10K tweets of test data and corresponding labels, with 10-fold cross-validation. Significantly, the obtained accuracy was very high (above 90%) and this is the reason why we did not investigate more advanced and recent classification methods. Our explanation for this excellent accuracy is that the dichotomic nature of the referendum fostered the emergence of sets of highly partisan hashtags, rarely used in a mix. Albeit the classifier uses the whole text of the tweets, it takes advantage of the presence of such hashtags to obtain remarkable performances. Unfortunately, we cannot guarantee equal accuracy of our classifier on other datasets – defining a high-quality and general purpose classifier being well beyond the scope of this paper.

On the whole, UNK tweets were substantially negligible – although this may be due to limitations of the classifier (Guarino et al. 2019) – while NO tweets were almost 1.5x more frequent than YES tweets, supporting the diffused belief that the NO front was significantly more active than its counterpart in the social debate. Significantly, we also obtained a continuous score in [-1,1] for users, since a user can be classified with the average score of his/her tweets. These user-level scores are used in the following sections for correlating polarization with other network properties of our corpus of users.

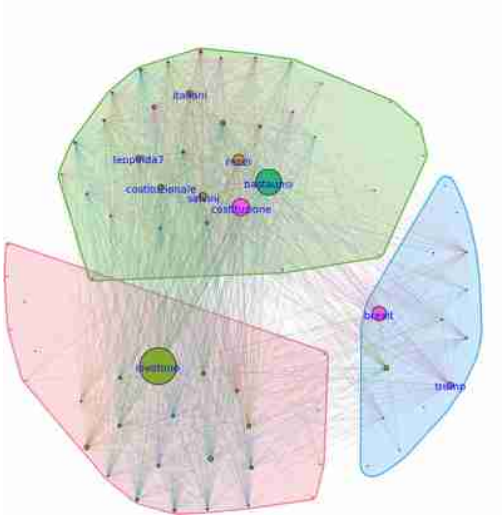
For the sake of clarity and completeness, the hashtag graph and its cluster-graph – wherein each cluster is contracted into a single node – are shown in Fig. 1. We see that: (i) hashtags used by the NO and YES supporters are strongly clustered; (ii) "neutral" hashtags (such as those used by international reporters) also cluster together; (iii) a few hashtags are surprisingly high-ranked, such as "#ottoemezzo", a popular political talk-show being central in the NO cluster – thus confirming regular patterns of behavior in the "second-screen" use of social network sites to comment television programs (Trilling 2015). In particular, the two largest clusters of hashtags clearly characterize the two sides: the YES cluster is dominated by the hashtags "#bastaunsì" ("a yes is enough") and "#iovotosi" ("I vote yes"), whereas the NO cluster by "#iovotono" ("I vote no"), "#iodicono" ("I say no") and "#renziacasa" ("Renzi go home"). In this perspective, the jargon of both communities show clear segregation and high levels of clustering by political alignments, as expected.

Polarized retweet graphs

The main objects of analysis of this paper are a set of interaction networks extracted from a Twitter dataset of more than 1.3 million tweets. Each of these networks is formally represented as a graph $G = (V, E)$, whose vertex set V models a corpus of social media users. Specifically, as often done in the literature [49, 34], we consider directed and weighted retweet graphs, wherein nodes are Twitter users and an edge e = (u, v) means that user u retweeted user v at least once in the considered corpus of tweets. In our graphs, edges



(a) Hashtag graph, with clusters highlighted. Vertex size is by pagerank and top 20 hashtags are annotated.



(b) Cluster graph. Vertex size is by cluster size and top 10 clusters are annotated with a reference hashtag.

Fig. 1 The hashtag graph and the associated cluster graph

are weighted by a parameter w_e equal to the number of retweets between a given pair of users. Nodes are instead endowed with a "polarization" attribute $p_u \in [-1, 1]$ – defined in the previous section – equal to the average polarization of the tweets and retweets of that user. Specifically, in this paper we consider the following six graphs:

- The **whole** retweet graph, obtained from the entire dataset.[2]
- The **P/D** (Propaganda/Disinformation) retweet graph, obtained from the set of all tweets that matched any of the queries defined in the "Data collection" section, i.e., tweets related to any of the 12 news stories.

[2]Precisely, we only consider the giant weakly connected component of this graph, which contains 92.55% of all vertices and 99.13% of all edges of the complete retweet graph.

- The **PI1**, **PI2**, **PI3** and **PI4** retweet graphs, induced by the set of tweets that satisfied each of the four selected propaganda items, taken individually.

The subgraph of the whole graph composed of the 1000 vertices having greatest pagerank is shown in Fig. 2. We can clearly see a few features of the graph that will be better discussed in the following: a general prevalence of NO edges (i.e., tweets), multiple NO-leaning clusters and a single main YES-leaning cluster.

A first relevant perspective on our dataset is obtained by considering how the vertex set of the P/D graph may be decomposed based on the belonging of its users to the individual PI graphs:

- 67.61% of all users in the P/D graph (i.e., 3666 users) are only involved in one of PI1, PI2, PI3, PI4;
- 16.17% of all users in the P/D graph (i.e., 877 users) are involved in two of PI1, PI2, PI3, PI4;
- 6.79% of all users in the P/D graph (i.e., 368 users) are involved in three of PI1, PI2, PI3, PI4;
- only 1.44% of all users in the P/D graph (i.e., 78 users) are involved in all four of PI1, PI2, PI3, PI4;
- 7.99% of all users in the P/D graph (i.e., 433 users) are involved in other PI besides PI1, PI2, PI3, PI4;

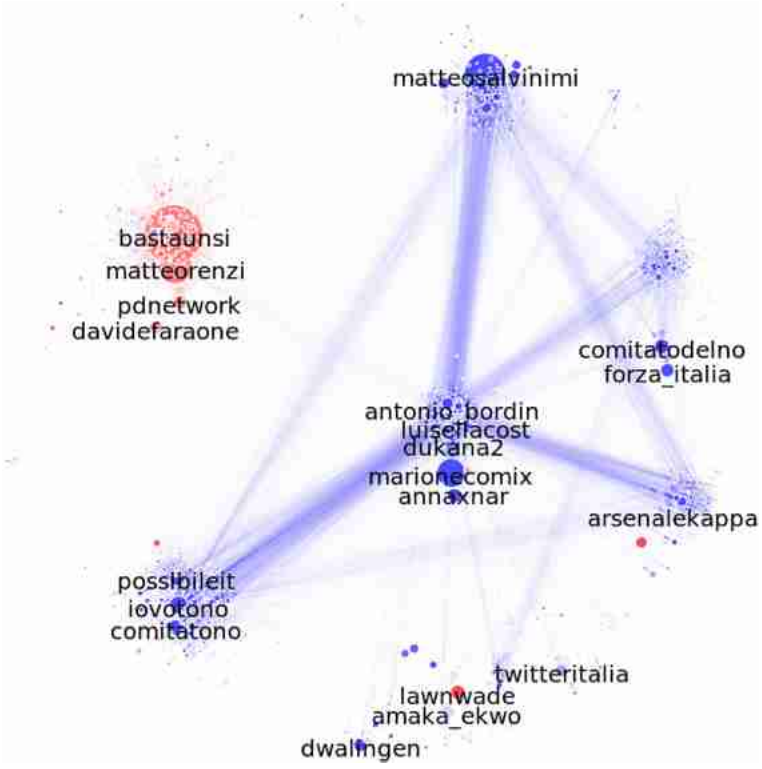


Fig. 2 The top 1000 vertices of the whole graph by pagerank. Vertex size is by pagerank, vertex and edge color is by polarization (YES=red, NO=blue), the 20 top users are annotated

Summing up, the four items of propaganda that we selected involve approximately 92% of all users of the P/D graph, with users only involved in other propaganda/fake news stories adding up to just 7.99%. We can thus safely focus on these four items without a significant loss in the generality of our results. At the same time, the fact that most users of the P/D graph were only involved in a single PI and that only a negligible fraction was involved in all four PIs warns us of the pitfalls of considering disinformation as a whole.

A second aspect to consider is the distribution of the polarization attribute p_u across the six graphs. For each of the six considered graphs, Fig. 3 shows the histogram and a kernel density estimate obtained considering the value of p_u for all users of the graph. Let us remind that $p_u \in [-1, +1]$ expresses the stance of user u with respect to the referendum in the range [NO,YES]. Overall, users appear to be strongly polarized, with two huge spikes at -1 and +1 for the whole graph. When we switch to networks of propaganda, however, users seem to be generally less polarized. This apparently counter-intuitive phenomenon is a consequence of our scoring method and of the much higher average activity of users involved in these networks. Indeed, a user's polarization is well-definite when that user has a single tweet and gets blurrier as the number of tweets increases, because of the contribution of many tweets not all of which are necessarily equally polarized. The average number of tweets per user in our propaganda networks is 8 to 14 times greater than the average computed over the whole graph. At the same time, users with a single tweet are 37% of the whole graph, but just 1% to 5% of the P/D and PIs graphs.

The distribution of PI1, PI2 and PI3 follows the overall trend of the P/D graph, that is, a general prevalence of NO users over YES users. Since all 4 selected items, as well as most of the 12 items, are pro-NO, this may be interpreted as a prevalence of propaganda over counter-propaganda. In that sense, PI4 is the exception: a clear example of a topic mostly used by one side (the YES coalition) to accuse the other of using deceptive propaganda. This is a first element in favour of the importance of accounting for polarization when characterizing these propaganda networks and their users.

Clustering structure

The clustering structure of a retweet graph highlights relevant properties of how users and groups of users interact with each other, and of how easily information flows through the graph. Along this line, recent work provided clear evidence that modularity-based clustering applied to retweet graphs brings to light communities of users with strong homophily/affiliation within which propaganda and polarized information spread especially well (Aragón et al. 2013; Becatti et al. 2019). By characterizing and comparing the clustering structure obtained for our six graphs through the well-known Louvain algorithm we expect to better understand the emergence of networks and sub-networks of propaganda and measure their persistence. To start, in Fig. 4 we show the size distribution of communities for the P/D and PIs retweet graphs: we rank the communities of each graph based on their size and we plot the size of each community on a log scale. At a high level, we see that the distributions of all PIs graphs are somewhat similar – especially for PI1 and PI3 – and that in all cases only a few clusters have a relevant size.

We now assess whether modularity based clustering detects communities of users with a clear attitude towards the referendum. To obtain a single polarization score for a given

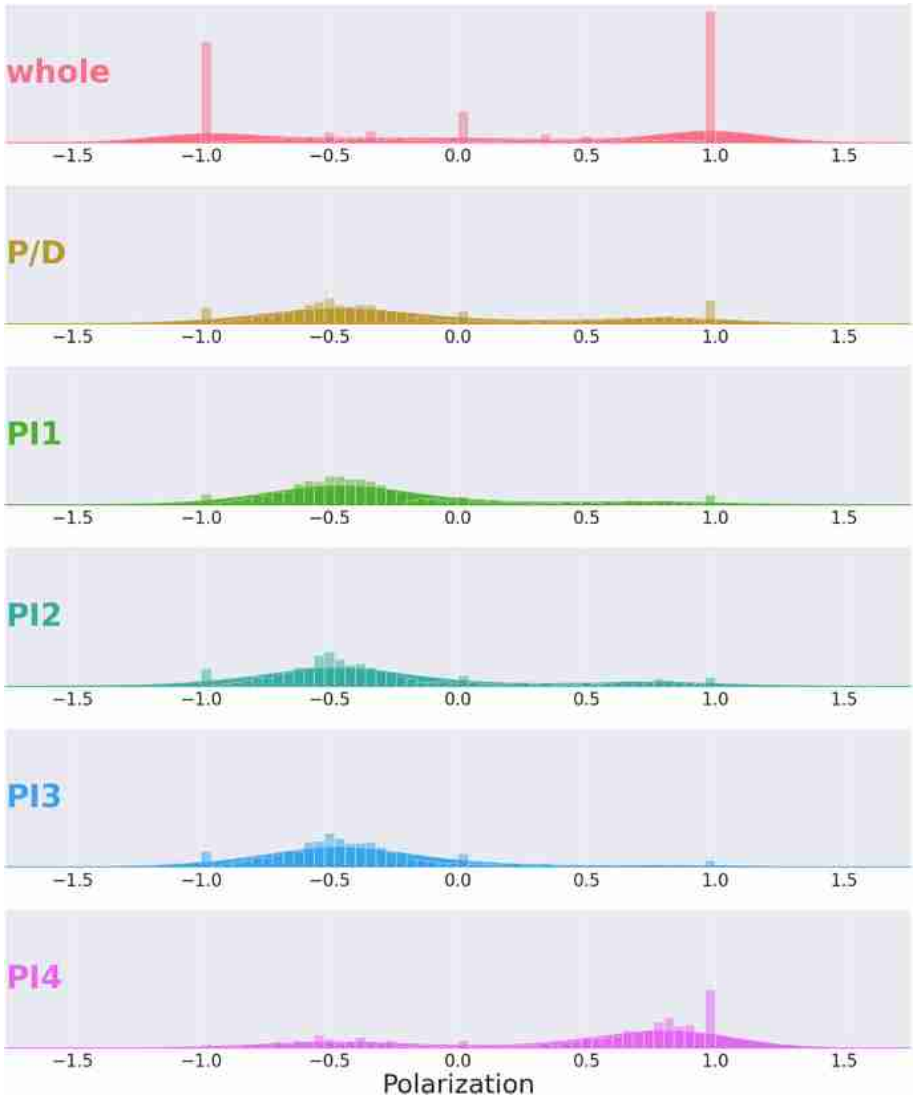


Fig. 3 Polarization of users involved in propaganda

cluster c, we computed the number of YES users in c, denoted Y_c, the number of NO users in c, N_c, and defined $p_c = \frac{Y_c - N_c}{Y_c + N_c}$. This definition guarantees that $p_c = +1$ if $N_c = 0$, $p_c = -1$ if $Y_c = 0$ and $p_c = 0$ if $Y_c = N_c$. Yet, if compared with just taking the average polarization of the users in c, this measure is more robust with respect to classification accuracy – under the assumption that telling apart YES and NO users is easier than measuring the exact polarization of each user. In Fig. 5 we consider the 10 largest clusters of each graph ranked by size and, for each of such clusters, we plot the polarization score p_c. The marker size is set proportional to the cluster size, whereas the marker color is also descriptive of the polarization in a range from blue (NO) to red (YES). We can clearly see that the clusters of the networks of propaganda are generally and significantly more polarized than the clusters of the whole graph. We also see that the overall prevalence of NO users in the P/D, PI1, PI2 and PI3 graphs already emerged in Fig. 3 is reflected in a greater number of NO clusters – the same happening in PI4 for the YES front.

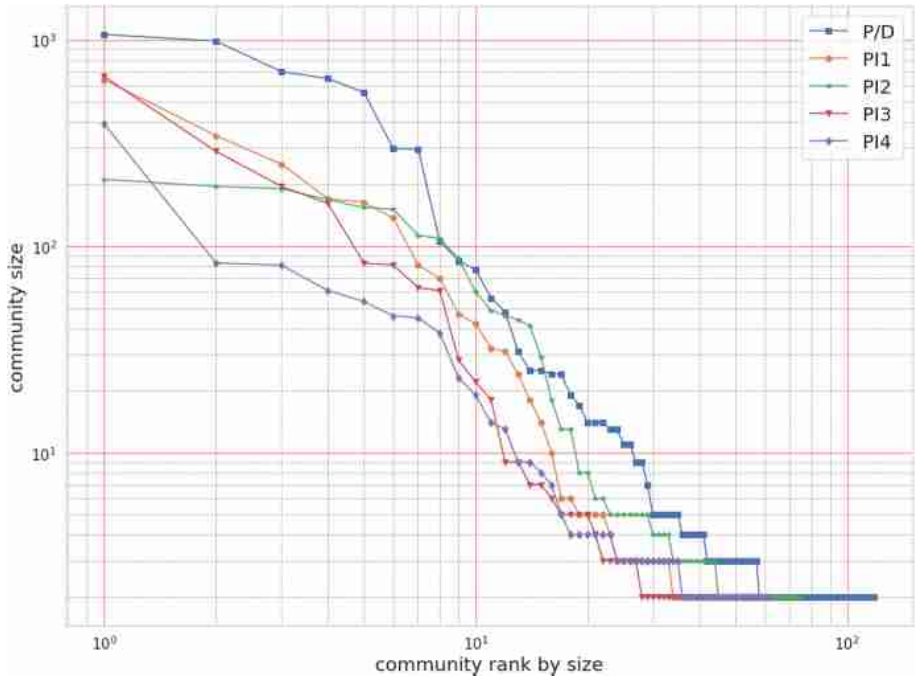


Fig. 4 Comparison of the clustering structure (cluster size distribution) for different propaganda graphs

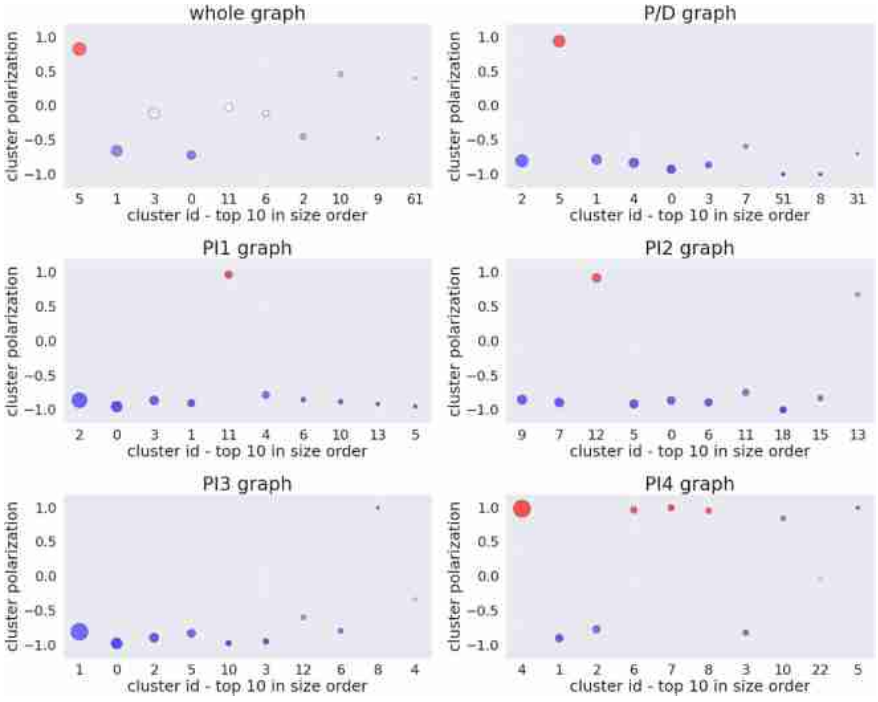


Fig. 5 The polarization of the top 10 clusters for each of the six graphs. The marker size is proportional to the cluster size. The polarization is also visible from the marker color

The main clusters of the whole graph deserve special attention. As already observed in the literature (Becatti et al. 2019), in fact, they quite clearly reflect political affiliation:

- Cluster 5 ($\approx$ 16K members) appears to group together members and supporters of the "Democratic Party", including Government members (such as PM Matteo Renzi and the Minister of Reforms Maria Elena Boschi), the official YES Committee and Renzi's foundation 'Leopolda', among the others.
- Custer 1 ($\approx$ 11K members) is expressive of the "5 Star Movement" community. Only two of the most active users (Minister Danilo Toninelli and Senator Elio Lannutti) are official party members, however, whereas the most influential actors belong to the militant base.
- Cluster 0 ($\approx$ 7.5K members) groups the members of the souverainist right, including the two politicians Matteo Salvini and Giorgia Meloni, their political parties, and a number of supporters.
- Cluster 2 ($\approx$ 3.5K members) clearly involves the "Forza Italia" members and advocates.

In this context, three large and barely-polarized clusters come to light. On the one hand, cluster 3 ($\approx$ 10K members) seems to validate the claim that "structure segregation and opinion polarization share no apparent causal relationship" (Prasetya and Murata 2020). It includes left-wing opponents to the referendum as well as several media accounts and has very low polarization (-0.04), a probable evidence of the willingness of the left-wing members of the NO alignment to maintain a cross-partisan interaction with the democrats. On the other hand, clusters 11 and 6 ($\approx$ 6K and $\approx$ 4K members, respectively) completely escape the party affiliation logic. Apart from @europeelects, which produces poll aggregation and election analysis in the European Union, we only found evidence of accounts belonging to international militants of the souverainist and anti-globalization movement: they are Brexit supporters, Italian pro-Trump advocates, or journalists covering such topics in their reporting activities.

Now, we aim at assessing to which extent the obtained clusters are influenced by the choice of a specific PI, that is, whether the patterns of cohesion among different users seem to be coherent across different topics of discussion. In Fig. 6 we use the Adjusted Mutual Information (AMI) to compare the clusters emerged in different graphs. Specifically, for each graph we draw a polyline showing the AMI between that graph's clustering and all other graphs' clustering. It is worth recalling that the AMI of two partitions is 1 if the two partitions are identical, it is 0 if the mutual information of the two partitions is the expected mutual information of two random partitions,[3] and it is negative if the mutual information of the two partitions is worse than the expected one. Of course, when comparing the partitions obtained for any two graphs, we just consider the users that are common to both graphs. In addition, in Fig. 7, we provide a more pointwise analysis of the 10 greatest communities of each PI graph, showing how users of these clusters distribute over the greatest 30 communities of the whole and P/D graphs. Precisely, in

[3] Here, the meaning of "random" depends on the choice of a distribution over the set of all possible partitions (Vinh et al. 2009)

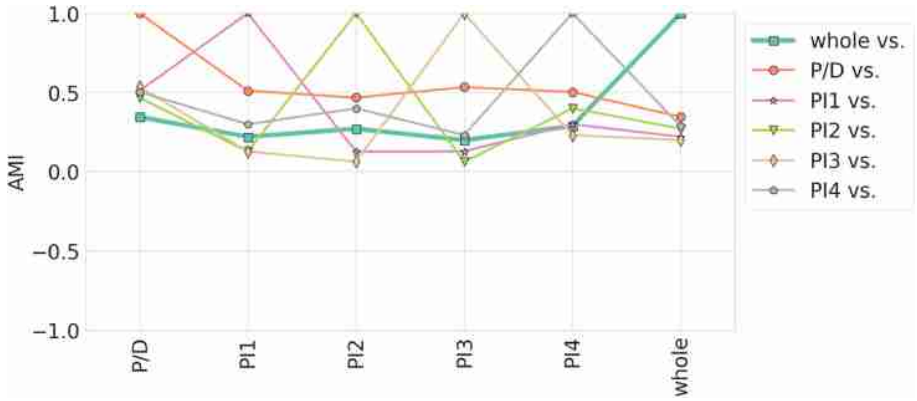


Fig. 6 Pairwise Adjusted Mutual Information of graphs' clusterings

each heatmap the cell at the intersection of row i and column j measures the proportion of users of cluster i in the considered PI graph that lie in cluster j of the compared graph.

The two figures together provide clear evidence that users are clustered in an rather unstable way, especially when we compare networks generated by individual PIs with the

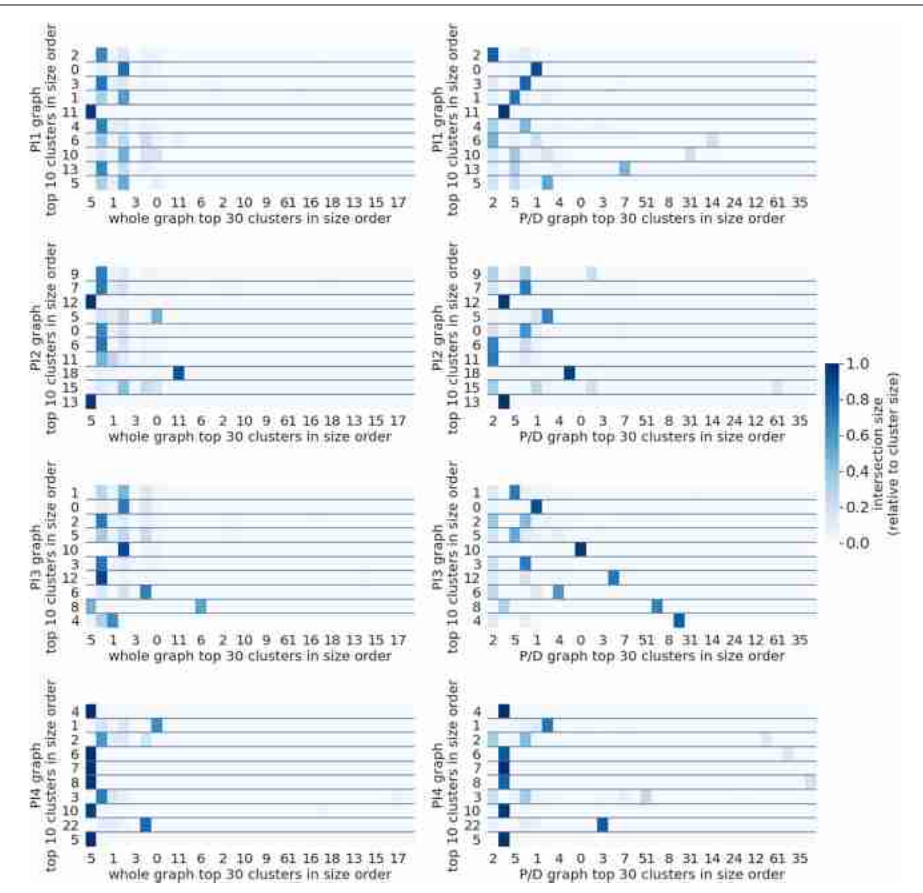


Fig. 7 Cluster-to-cluster intersection size: top 10 clusters of each PI graph *vs.* top 30 clusters of the whole retweet graph and the P/D graph clusters. The intersection size is normalized by the considered PI cluster size (i.e., each row is individually normalized and sums up to (almost) 1)

whole retweet graph and with each other. The topological organization of the NO front is adequately expressive of different ideological affiliations of NO sponsors, but these differences are not clearly visible in the participation to clusters of the PI1, PI2 and PI3 networks. Assuming that selective exposure and social validation are core driving polarization mechanisms (Prasetya and Murata 2020), two main interpretations are possible: either (i) being generally NO-leaning is enough to trigger the exposure to these three PIs, with the actual political community a user belongs to playing a marginal role; or (ii) the interactions occurring globally on Twitter – and, as such, global information flows – are only partially responsible of the tendency of users to diffuse propaganda and disinformation items. In PI4, on the other hand, most clusters are *de facto* sub-clusters of a macro-community of the whole retweet graph. This is easily explained by the different polarization emerged in Fig. 5: the macro-community is cluster 5, which we already identified as the "YES community". The YES cluster seems to be driven by both an effort of community building and the attempt to de-legitimate the NO front by debunking its news-claims and propaganda items. As a consequence, YES users stay attached in the P/D and in other PI graphs, while they splits into sub-communities in PI4, showing a stronger degree of internal homogeneity and highlighting a polarized conversational archetype, with partisan actors and segregated community structure and discussion.

Users' centrality

In this section we study the role played by the users in the four propaganda items selected and discussed in the previous sections. To assess the activity of each user we compute the following centrality measures on the retweet graphs: PageRank, In-Degree, Out-Degree, Authority Score and Hub Score (Kleinberg 1999). The centrality measures we chose are often used for networks analysis and their interpretation depends on the phenomenon modeled by the network. In our graph, the In-Degree tells us which are the users that are more often retweeted, i.e., the users creating contents that are spread on the network. The PageRank tells us which are the users that are most likely "visited", i.e., the users whose contents are most probably read if the retweets graph is used to surf the network. The Out-Degree tells us which are the users that more often retweet, i.e., the users playing a main role in the information diffusion. Finally the Hub Score and Authority Score are interconnected: the former tells us which are the users that more often retweet contents created by an *authority*, we call these users *hubs*; the latter tells us which are the users creating the main contents about a discussion topic, we call these users *authorities*. The main difference between the Hub Score and the Out-Degree is about the content of the retweets done by a user, in the former case the user retweets authoritative contents, in the latter case the user does not show any preference about the tweet's origin. Similarly, the main difference between the Authority Score and the In-Degree resided in the type of users that usually retweet contents produced by a given user, in the former case the users retweeting these contents are *hubs*, in the latter case there is no distinction among users. Thus, a user has a high Authority Score if she has a high degree and the users retweeting her contents are hubs. Tables 1, 2, 3, 4 and 5 show for each propaganda item the top 10 users of each metric. The color of each cell denotes the polarization of the user, blue is used for NO supporters and red is used for YES supporters. The color's intensity shows how strong is the polarization, i.e., the darker is the color the more polarized is the user.

Table 1 Whole retweet graph: top 10 accounts by centrality

In-Degree	Out-Degree	Authority Score	Hub Score	PageRank
matteosalvinimi	iovotono	antonio_bordin	marino29b	bastaunsi
antonio_bordin	marino29b	marionecomix	iovotono	matteosalvinimi
bastaunsi	gincarbone	dukana2	franco_dimuro	matteorenzi
marionecomix	franco_dimuro	iovotono	nativiitaliani	marionecomix
dukana2	nativiitaliani	sevenseasmarina	luisaloffredo28	antonio_bordin
iovotono	gjscco	andfranchini	demian_yexil	iovotono
matteorenzi	luisaloffredo28	beatricedimadi	gincarbone	possibileit
claudiodeglinn2	lelloesposito5	annaxnar	il_brigante07	comitatono
comitatono	demian_yexil	oinot49	gjscco	comitatodelno
sevenseasmarina	giorgiomorresi	cremaschig	giorgiomorresi	dukana2

Our results show a few relevant aspects. First, if we look at the whole retweet network the most active users are almost all NO supporters, as showed in Table 1, albeit the number of NO and YES users is quite balanced in the network – as showed in Fig. 3. Indeed, we find only few accounts belonging to the YES supporters in the top position of the five metrics. Additionally, by analyzing individual propaganda items we observe that the set of most active users, their ranking and their polarization change depending on the considered PI and metrics. In PI1, PI2 and PI3 the most active accounts are NO supporters, while in PI4 the YES supporters are the most active and numerous, in accordance with Fig. 3 and despite PI4 also being a pro-NO item. The analysis of users polarization is thus essential to understand the role played by the main actors inside each network: if we only considered the centrality of the users, without looking at their polarization, we would not be able to distinguish between accounts that are contributing to the diffusion of a fake news and accounts that are working against, i.e., the debunkers. We also see, again in accordance with the analysis presented in the "Clustering structure" section, that different PIs see different users take on different roles. Well known public figures – such as "mattosalvinimi", "giorgiameloni" and "matteorenzi" – are a minority with respect to grassroots activists, and users playing a central role in a specific network of propaganda and/or with respect to a specific metrics are absent or not as relevant in other cases – such as "cinmir89" or "proudman811".

To further investigate the persistence of these rankings across different networks of propaganda, in Fig. 8 we present a set of correlation matrices that broadly corroborate the

Table 2 PI1 retweet graph: top 10 accounts by centrality

In-Degree	Out-Degree	Authority Score	Hub Score	PageRank
antonio_bordin	nativiitaliani	antonio_bordin	nativiitaliani	antonio_bordin
matteosalvinimi	marino29b	matteosalvinimi	giorgiomorresi	didimiero
dukana2	il_brigante07	dukana2	cinmir89	matteosalvinimi
francotrax	celestinoceles7	francotrax	andreazanettin	francotrax
carloalterego	proudman811	carloalterego	il_brigante07	dukana2
claudiodeglinn2	lelloesposito5	eliolannutti	cretellaroberta	eliolannutti
eliolannutti	marobe997	newsinunclick	soloio0509	penelopy2000
5bc32772e3fb467	franco_dimuro	possidonio_gg	archidevivaio	adrimcmlxi
patriziarametta	giorgiomorresi	5bc32772e3fb467	dopiot	ipredicatore
ipredicatore	kirumakataossi1	claudiodeglinn2	marino29b	toscaross

Table 3 PI2 retweet graph: top 10 accounts by centrality

In-Degree	Out-Degree	Authority Score	Hub Score	PageRank
dukana2	iovotono	dukana2	marino29b	comitatodelno
pdnetwork	marino29b	sevenseasmarina	nativiitaliani	renatobrunetta
comitatodelno	nativiitaliani	antonio_bordin	proudman811	sevenseasmarina
sevenseasmarina	il_brigante07	ermannokilgore	iovotono	dukana2
antonio_bordin	uleprr	fmcastaldo	cretellaroberta	pdnetwork
advalita	gincarbone	comitatodelno	gjscco	antonio_bordin
ermannokilgore	battistabd	rossellafidanza	uleprr	advalita
fmcastaldo	franco_dimuro	deboramau	il_brigante07	ermannokilgore
rossellafidanza	mad13021966	advalita	battistabd	inarratore
deboramau	proudman811	annaxnar	cocchi2a	fmcastaldo

previous findings. Specifically, for each centrality measure we report the pairwise correlation between the rankings produced by that measure on different graphs, in order to better understand the role of the users that were active in more than one propaganda item. We rely on Spearman's rank correlation coefficient, rather than the widely used Pearson's, because we are neither especially interested in verifying linear dependence, nor we do expect to find it. We are more interested in the possible monotonic relationship between centrality measures as determined by Spearman's correlation.

As already observed, combining the centrality and polarization data, we notice that in the PI4 network there is a different community of users that is active and that is spreading information with respect to the other PI networks. This behaviour is clearly visible from the Hub Score matrix (Fig. 7d) and partially from the Out-Degree matrix (Fig. 7b). In the former the anti-correlation in row PI4 shows the existence of a different community of spreaders in PI4 with respect to other PI. A community composed of users that are absent, less active or play a different role in other networks. In the Out-Degree matrix we have almost no correlation in row PI4 except for PI2. This difference is due to the presence of a small, non-negligible, community of YES supporters in PI2 as showed in Fig. 5. Whereas, the PageRank and In-Degree matrices show that the relevance of the accounts creating contents is more stable than those diffusing the information. Finally, the Authority Score matrix shows that, although the contents creator accounts are stable, their role change in the network. The same account is considered more authoritative in one network and less in the other.

Table 4 PI3 retweet graph: top 10 accounts by centrality

In-Degree	Out-Degree	Authority Score	Hub Score	PageRank
claudiodeglinn2	marino29b	claudiodeglinn2	nativiitaliani	claudiodeglinn2
matteosalvinimi	gincarbone	matteosalvinimi	giorgiomorresi	angelosica1965
luisaloffredo28	nativiitaliani	patriotail	mania48mania53	matteosalvinimi
patriotail	luisaloffredo28	luisaloffredo28	gincarbone	sevenseasmarina
sevenseasmarina	giorgiomorresi	deglclaudio	caspanistefania	deglclaudio
giorgiameloni	malaspinadavide	civico21	pietrof70	patriotail
liberatilinda	piras_zia	angelosica1965	ilpellicano88	xmeridio78
civico21	ori254	sevenseasmarina	celestinoceles7	civico21
deglclaudio	celestinoceles7	carmentpf	archidevivaio	liberatilinda
5bc32772e3fb467	claudiodeglinn2	valy_s	gidal_randagio	toscaross

Table 5 PI4 retweet graph: top 10 accounts by centrality

In-Degree	Out-Degree	Authority Score	Hub Score	PageRank
bastaunsi	danieledvpd	bastaunsi	danieledvpd	bastaunsi
fnicodemo	angelinascanu	fnicodemo	rtgovernorenzi	renatobrunetta
renatobrunetta	amtomarchio	thelambkin_	giordanobattini	ilmattinale
thelambkin_	rtgovernorenzi	magdazanonii	alcinx	fnicodemo
magdazanonii	italiarecord	piercamillo	albertoforesti3	fi_online_
paolocristallo	lcungi	belpassijessica	italiarecord	renatapolverini
eugeniocardi	mursino71	serracchiani	amtomarchio	thelambkin_
piercamillo	alcinx	eugeniocardi	alfuturosi	magdazanonii
arsenalekappa	albertoforesti3	diegozardini	angelinascanu	comitatodelno
ilmattinale	giordanobattini	unitaonline	ruiciccio	paolocristallo

What is happening in the other networks can be better understood by looking at Fig. 7f-i where we computed separately the correlation for NO and YES supporters for the Authority and Hub Scores, that overall appear to be the most informative. To better focus our analysis we excluded the PI4 row. Our results show that among YES supporters the content creators accounts are not stable and their role change depending on the propaganda item selected. On the other hand, the role of the accounts spreading information is more stable, meaning that for different networks there are different authorities, but the hubs are the same. For what concerns the NO supporters we have that both authorities and hubs

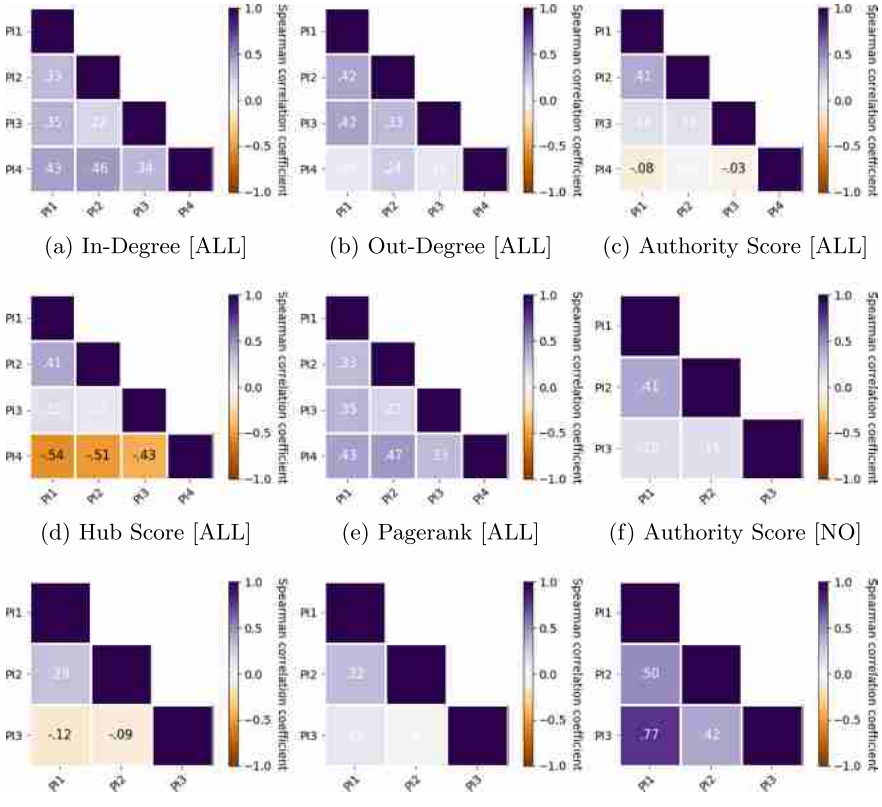


Fig. 8 Spearman's rank correlation coefficients

relevance changes depending on the propaganda network. Thus there is probably a more efficient synergy among NO supporters between authority and hub accounts.

Conclusions

The paper aimed at providing new insights into the dynamics of propaganda networks on Twitter. The results of our study are partly in line with existing research. Modularity-based clustering, applied to retweet graphs, pictured a wide panorama of communities of users with strong homophily/affiliation and polarized position. As expected, the clusters of propaganda networks were generally and significantly more polarized than the clusters of the whole graph and the topological organization proved to be highly representative of the ideological affiliation of users. The comparison between clusters in different graphs reveals that users' clusters are rather dynamic, particularly when comparing networks generated by individual propaganda items with the whole retweet graph and with each other. It seems that global clusters, often associated with information exposure, are only partially responsible of the tendency of users to diffuse propaganda and disinformation items. When it comes to taking a position on a controversial topic, users tend to group with different people with respect to those they usually connect to in the whole graph, and the "high-level" polarization of a user – such as the NO *vs.* YES leaning in our case – may have a more prominent role than his/her political affiliation. This is especially visible for users involved in propaganda – as opposed to counter-propaganda.

The combined analysis of cluster-to-cluster intersections and centrality metrics additionally indicates how different propaganda items are associated to different users with authoritative roles. The correlation of centrality metrics across different networks provides further insights: (i) the Authority and Hub Score seem the most informative metrics for studying networks of propaganda, thanks to their ability to tell apart content creators and spreaders; (ii) the role of content creators is taken by different users for different propaganda items, independently of clusters polarization; (iii) spreaders are instead generally more "consistent". Overall, the propaganda community depicted in this study, far from being monolithic, has a considerable degree of internal variability, in terms of central actors, topics and opinion polarization. Polarization with respect to a main theme, transversal to the considered propaganda items, emerged as a fundamental parameter in governing users behavior. A side result of the present paper is the identification of a few expedients and precautions to be used in practice. For instance, we showed that the authority and hub scores unveil different players of a propaganda network, and that real-time detection of propaganda and disinformation campaigns must be built on top of a reliable polarization measure. To this end, it must be kept in mind that users' polarization (on a specific issue) and political partisanship do not always coincide: we showed that the topic of debate may significantly alter the community structure of an interaction network, and thus the perceived affiliation of its users. Further directions of research could involve other clustering algorithms as well as dynamic influence metrics, in order to gain deeper knowledge on the relationship between exposure to propaganda and the general structure of users interaction.

Another issue we explicitly chose not to cover involves the determinants of user centrality in a debate (why or how a user gained a central role?), nor to detect coordinated bot attacks that possibly boosted the centrality of a Twitter profile. We rather focus on the *perceived* centrality of a user, regardless of what caused it, to show that: (i) the centrality itself

is of limited use if not accompanied with a polarization analysis, e.g., to distinguish propaganda from counter-propaganda/debunking; (ii) using different metrics make it possible to detect different roles in the network, and such roles vary from one disinformation item to another. That said, the analysis highlighted evidence of a major coordination effort in the NO front, which is where the considered propaganda and disinformation items were more prevalent. Understanding whether this coordination was supported by bots is left to future work.

Abbreviations
AMI: Adjusted Mutual Information; P/D: Propaganda/Disinformation; PI: Propaganda Item; PM: Prime Minister; UNK: Unknown

Acknowledgements
Not applicable.

Authors' contributions
SG, NT, ACe and ACh designed the study. ACh acquired the data. SG and ACe created the software used to perform the data analysis. SG, NT and ACe interpreted the results and wrote the paper. All author(s) revised the work and read and approved the final manuscript.

Funding
This work was supported in part by the Project "SOMA", funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 825469. The European Commission had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. Any opinion, finding, and conclusions expressed in this paper only reflect the views of the authors.

Availability of data and materials
Part of the code used in this paper will be included in the network analysis toolbox DisInfoNet, currently under development by the partners of the Project "SOMA" at https://gitlab.com/s.guarino/disinfonet. DisInfoNet is presented in a previous conference paper (Guarino et al. 2019) and will be released by the end of the SOMA Project. The entire dataset used during the current study is not publicly available due to Twitter's policies. The ids of the tweets are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Institute for Applied Mathematics, National Research Council, Rome, Italy. [2]Luiss "Guido Carli" University, Rome, Italy. [3]Linkalab, Cagliari, Italy.

Received: 28 February 2020 Accepted: 17 July 2020
Published online: 04 September 2020

References
Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–36
Alonso-Muñoz L, Casero-Ripollés A (2018) Communication of european populist leaders on twitter: Agenda setting and the 'more is less' effect. El profesional de la información 27(6):1193–02
Aragón P, Kappler KE, Kaltenbrunner A, Laniado D, Volkovich Y (2013) Communication dynamics in twitter during political campaigns: The case of the 2011 spanish national election. Policy Internet 5(2):183–206
Bakker TP, De Vreese CH (2011) Good news for the future? young people, internet use, and political participation. Commun Res 38(4):451–470
Becatti C, Caldarelli G, Lambiotte R, Saracco F (2019) Extracting significant signal of news consumption from social networks: the case of twitter in italian political elections. Palgrave Commun 5(1):1–16
Bessi A, Ferrara E (2016) Social bots distort the 2016 us presidential election online discussion. First Monday 21(11–7)
Bovet A, Makse HA (2019) Influence of fake news in twitter during the 2016 us presidential election. Nat Commun 10(1):7
Brody DC, Meier DM (2018) How to model fake news. arXiv preprint arXiv:1809.00964
Burckhardt P, Duch R, Matsuo A (2016) Tweet as a tool for election forecast: UK 2015. General election as an example. [online]. http://asiapolmeth.princeton.edu/sites/default/files/polmeth/files/uk_election_tweets_asia_polmeth.pdf
Castanho Silva B, Vegetti F, Littvay L (2017) The elite is up to something: Exploring the relation between populism and belief in conspiracy theories. Swiss Polit Sci Rev 23(4):423–443
Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th International Conference on World Wide Web. ACM, New York. pp 675–684
Ceron A, Curini L, Iacus SM (2016) Politics and big data: nowcasting and forecasting elections with social media. Taylor & Francis
Ciampaglia GL, Shiralkar P, Rocha LM, Bollen J, Menczer F, Flammini A (2015) Computational fact checking from knowledge networks. PloS ONE 10(6):0128193

Conover M, Ratkiewicz J, Francisco MR, Gonçalves B, Menczer F, Flammini A (2011) Political polarization on twitter. Icwsm 133:89–96
Crowell C (2017) Our approach to bots & misinformation. Twitter public policy
Del Vicario M, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W (2017) Modeling confirmation bias and polarization. Sci Rep 7:40391
Feng S, Banerjee R, Choi Y (2012) Syntactic stylometry for deception detection. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Association for Computational Linguistics, Jeju Island. pp 171–175
Feng VW, Hirst G (2013) Detecting deceptive opinions with profile compatibility. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing. Asian Federation of Natural Language Processing, Nagoya. pp 338–346
Gayo-Avello D, Metaxas PT, Mustafaraj E (2011) Limits of electoral predictions using twitter. In: Fifth International AAAI Conference on Weblogs and Social Media. The AAAI Press, Menlo Park, California
Guarino S, Trino N, Chessa A, Riotta G (2019) Beyond fact-checking: Network analysis tools for monitoring disinformation in social media. In: International Conference on Complex Networks and Their Applications. Springer, Lisbon. pp 436–447
Guess A, Nagler J, Tucker J (2019) Less than you think: Prevalence and predictors of fake news dissemination on facebook. Sci Adv 5(1):4586
Higgins K (2016) Post-truth: a guide for the perplexed. Nat News 540(7631):9
Karadzhov G, Nakov P, Màrquez L, Barron-Cedeno A, Koychev I (2017) Fully automated fact checking using external sources. arXiv preprint arXiv:1710.00341
Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. J ACM (JACM) 46(5):604–632
Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Metzger MJ, Nyhan B, Pennycook G, Rothschild D, et al (2018) The science of fake news. Science 359(6380):1094–1096
Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: International Conference on Machine Learning, vol. 32. JMLR: W&CP, Beijing. pp 1188–1196
Lévy P (2002) Cyberdémocratie: essai de philosophie politique. In: A Inteligência Coletiva. Odile Jacob, Paris
Löfstedt R (2005) Risk management in post-trust societies. Springer, New York: Palgrave Macmillan
Margolin DB, Hannak A, Weber I (2018) Political fact-checking on twitter: when do corrections have an effect?. Polit Commun 35(2):196–219
Markowitz DM, Hancock JT (2014) Linguistic traces of a scientific fraud: The case of diederik stapel. PloS ONE 9(8):105937
Mastinu L (2016) TOP 10 Bufale e disinformazione sul Referendum. www.bufale.net/top-10-bufale-e-disinformazione-sul-referendum/. Accessed 05 July 2019
Nichols T (2017) The death of expertise: The campaign against established knowledge and why it matters. Wiley Online Library
Papacharissi Z, de Fatima Oliveira M (2012) Affective news and networked publics: The rhythms of news storytelling on# egypt. J Commun 62(2):266–282
Pariser E (2011) The filter bubble: what the internet is hiding from you. Penguin UK
Pierri F, Artoni A, Ceri S (2020) Investigating italian disinformation spreading on twitter in the context of 2019 european elections. PloS ONE 15(1):0227821
Politica RP (2016) La notizia più condivisa sul referendum? È una bufala. https://pagellapolitica.it/blog/show/148/la-notizia-pi%C3%B9-condivisa-sul-referendum-%C3%A8-una-bufala. Accessed 05 July 2019
Post RI (2016) Nove bufale sul referendum. www.ilpost.it/2016/12/02/bufale-referendum/. Accessed 05 July 2019
Prasetya HA, Murata T (2020) A model of opinion and propagation structure polarization in social media. Comput Soc Networks 7(1):1–35
Qiu X, Oliveira DF, Shirazi AS, Flammini A, Menczer F (2017) Limited individual attention and online virality of low-quality information. Nat Hum Behav 1(7):0132
Schroeder R (2018) Digital media and the rise of right-wing populism. Soc Theory Internet Media Technol Glob:60–81
Shao C, Ciampaglia GL, Flammini A, Menczer F (2016) Hoaxy: A platform for tracking online misinformation. In: Proceedings of the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, Montréal. pp 745–750
Shao C, Ciampaglia GL, Varol O, Yang K-C, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. Nat Commun 9(1):4787
Shao C, Hui P-M, Wang L, Jiang X, Flammini A, Menczer F, Ciampaglia GL (2018) Anatomy of an online misinformation network. PloS ONE 13(4):0196087
Shin J, Thorson K (2017) Partisan selective sharing: The biased diffusion of fact-checking messages on social media. J Commun 67(2):233–255
Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: A data mining perspective. ACM SIGKDD Explor Newsl 19(1):22–36
Silverman C, Singer-Vine J (2016) Most americans who see fake news believe it, new survey says. BuzzFeed News 6. https://www.buzzfeednews.com/article/craigsilverman/fake-news-survey
Skurnik I, Yoon C, Park DC, Schwarz N (2005) How warnings about false claims become recommendations. J Consum Res 31(4):713–724
Subrahmanian V, Azaria A, Durst S, Kagan V, Galstyan A, Lerman K, Zhu L, Ferrara E, Flammini A, Menczer F, et al (2016) The darpa twitter bot challenge. arXiv preprint arXiv:1601.05140
Sunstein CR (2001) Republic.com. Princeton university press
Trilling D (2015) Two different debates? investigating the relationship between a political debate on tv and simultaneous comments on twitter. Soc Sci Comput Rev 33(3):259–276. https://doi.org/10.1177/0894439314537886
Tucker J, Guess A, Barberá P, Vaccari C, Siegel A, Sanovich S, Stukal D, Nyhan B (2018) Social media, political polarization, and political disinformation: A review of the scientific literature. Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)

Van der Linden S, Leiserowitz A, Rosenthal S, Maibach E (2017) Inoculating the public against misinformation about climate change. Glob Challenges 1(2):1600008
Vicario MD, Quattrociocchi W, Scala A, Zollo F (2019) Polarization and fake news: Early warning of potential misinformation targets. ACM Trans Web (TWEB) 13(2):10
Vinh NX, Epps J, Bailey J (2009) Information theoretic measures for clusterings comparison: is a correction for chance necessary? In: Proceedings of the 26th Annual International Conference on Machine Learning. Association for Computing Machinery (ACM), Montréal. pp 1073–1080
Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151
Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R (2018) Detection and resolution of rumours in social media: A survey. ACM Comput Surv (CSUR) 51(2):32

Publisher's Note