ARTICLE

Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections

Carolina Becatti[1], Guido Caldarelli[1,2,3,4], Renaud Lambiotte[5] & Fabio Saracco [1]

ABSTRACT According to the Eurobarometer report about EU media use of May 2018, the number of European citizens who consult on-line social networks for accessing information is considerably increasing. In this work we analyse approximately 10^6 tweets exchanged during the last Italian elections held on March 4, 2018. Using an entropy-based null model discounting the activity of the users, we first identify potential political alliances within the group of verified accounts: if two verified users are retweeted more than expected by the non-verified ones, they are likely to be related. Then, we derive the users' affiliation to a coalition measuring the polarisation of unverified accounts. Finally, we study the bipartite directed representation of the tweets and retweets network, in which tweets and users are collected on the two layers. Users with the highest out-degree identify the most popular ones, whereas highest out-degree posts are the most "viral". We identify significant content spreaders with a procedure that allows to statistically validate the connections that cannot be explained by users' tweeting activity and posts' virality, using an entropy-based null model as benchmark. The analysis of the directed network of validated retweets reveals signals of the alliances formed after the elections, highlighting commonalities of interests before the event of the national elections.

[1] IMT School for Advanced Studies, P.zza S. Francesco 19, 55100 Lucca, Italy. [2] Istituto dei Sistemi Complessi (CNR) UoS Università "Sapienza", P.le A. Moro 2, 00185 Rome, Italy. [3] ECLT, Università di Venezia "Ca' Foscari", S. Marco 2940, 30124 Venice, Italy. [4] Catchy srl, Talent Garden Poste Italiane, Via Giuseppe Andreoli 9, 00195 Rome, Italy. [5] Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, UK. Correspondence and requests for materials should be addressed to C.B. (email: carolina.becatti@imtlucca.it)

Introduction

In recent years, the computer revolution we are witnessing made evident that the interconnection between individuals in the society formed a (non regular) lattice on which news and gossips propagate from one person to another one (Barabási, 2011). Indeed, the way Europeans approach news and information has drastically changed. According to the Eurobarometer report about EU media use of May 2018 (TNS opinion social and Directorate-General Communications, 2018), printed press is consulted everyday by 28% of the EU citizens, following a decreasing trend, while is never consulted by approximately 20% of them. On the other hand, the daily access to the Internet increased up to 65% of the population (starting from a 45% in 2010) and the increase in the percentage of citizens using daily on-line social network for accessing information is even more striking, changing from 18% in 2010 to 42% in 2017. This movement towards new technologies is especially true for Italy: according to the report (AGCOM, 2018) the Italians' most consulted medium to gather information in 2017 is the TV, followed by the Internet, the radio and the newspapers (respectively, 68.8, 41.8, 24.6% and 17.3% of the population daily access these channels). Despite the informative power of the on-line sources seems to be growing, it is paradoxically in opposition to a general distrust towards new media: in Italy of 2017, the Internet (intended as on-line social networks, blogs, news portals, on-line newspapers websites, etc.) is perceived as the less reliable information source, while TV is seen as broadcasting the most trustworthy news.

The different ways people access and consume information have always been a topic of interest for the social sciences especially during electoral events, when the research is focused on the identification of the determinants of citizens' political preferences. Preliminary works (Lazarsfeld et al., 1944; Katz and Lazarsfeld, 1955; Berelson et al., 1954) were fundamental in understanding how one's political choice is affected by interpersonal networks of different nature (such as friendship, colleagues, family). This idea is even more relevant nowadays, when the citizen's opinion is not only related to her/his *real-life* social sphere but also to the *virtual* circle of connections, who express their beliefs on different social platforms. Therefore, also in more recent years, different analyses based on complex network theory (Newman, 2010; Caldarelli, 2010) have been undertaken in order to tackle different aspects of these phenomena.

Due to their increasing popularity and usage for accessing information, several studies are based on the analysis of on-line users' behaviour in conjunction with political events, such as: the raising of grass-root democratic protests as Arab springs (González-Bailón et al., 2011), Occupy Wall Street (Conover et al., 2011) or the Spanish "Indignados" (González-Bailón et al., 2013); the electoral campaign itself, that has been focused time by time on USA (Adamic and Glance, 2005; Diakopoulos and Shamma, 2010; DiGrazia et al., 2013; Bekafigo and McBride, 2013; Badawy et al., 2018; Bovet et al., 2018; Bovet and Makse, 2018), Australia (Gibson and McAllister, 2006; Bruns and Stieglitz, 2012), Norway (Enli and Skogerbø, 2013), Spain (Borondo et al., 2012; Stella et al., 2018b), Italy (Caldarelli et al., 2014; Del Vicario et al., 2017a; Bindi et al., 2018; Stella et al., 2018a), France (Ferrara, 2017) and UK (Cram et al., 2017; Del Vicario et al., 2017b).

One of the consequences of the use of the Internet and on-line social networks for accessing information is a general shift from mediated to disintermediated news consumption, because of the fact that also sources that are unrelated to the standard information channels can participate to news spreading (AGCOM, 2018). This fact has led to a range of documented phenomena: users tend to focus on information reinforcing their opinion (*confirmation bias* (Quattrociocchi et al., 2014; Del Vicario et al.,

2016b, 2016a; Schmidt et al., 2018a, 2018b) and to group in clusters of people with similar viewpoints, forming the so called *echo chambers*, a phenomenon that is largely studied in on-line settings (Bakshy et al., 2015; Nikolov et al., 2015; Del Vicario et al. 2016b, 2016a; Schmidt et al., 2018a, 2018b) but that also have conspicuous repercussions offline (Hampton et al., 2017; Di Fonzo, 2011). Remarkably, even if it is becoming of public interest due to the diffusion of the political debate in on-line social media, the reinforcement of personal opinion due to the influence of reference groups was already analysed in previous studies (Oshagan, 1996; Moy et al., 2001). The different dynamics that the public debate follows on social-network platforms is also remarkable: the time evolution of viral non-verified contents is more persistent than the verified equivalent (Del Vicario et al., 2016a) and "negative" messages spread faster than "positive" ones, even if the latter reach on average a wider audience (Ferrara and Yang, 2015). Moreover, the analysis of time evolution of the activities in social platforms helps to predict the trend of retweets (Kobayashi and Lambiotte, 2016), the interactions of a single user with her/his neighbours (Tabourier et al., 2016) and to detect future developments of information campaign at an early stage (Varol et al., 2017) or "astroturf" campaigns (Ferrara, 2017).

The analysis of users' on-line behaviour during political event is the major topic of the present work too. However, we observe that the study of on-line social networks is often complicated by their complex, intertwined organisation and their strong heterogeneity. Therefore, with respect to the previous literature we approach this topic on a different manner, relying on carefully constructed null models to identify the network's salient features as significant deviations from them. An unbiased entropy-based null-model can be obtained following the information theory derivation of statistical physics (Jaynes, 1957; Squartini Tiziano and Garlaschelli Diego, 2017; Cimini et al., 2018): starting from an observed network, define an ensemble comprehending all possible networks with the same amount of nodes[1] but variable number of links, passing from an empty network to the fully connected one, via all possible link configurations. Then define a probability distribution over the ensemble: the shape of this distribution can be derived through the entropy maximisation under certain constraints (Park and Newman, 2004), i.e., preserving the average values over the ensemble of some quantities of interest. In order to obtain an estimate for the probability distribution parameters, we maximise the likelihood to observe the real network (Garlaschelli and Loffredo, 2008; Squartini and Garlaschelli, 2011). The crucial point of the above construction is the role of the constraints: in order to provide a reliable null-model, constraints should represent important properties of the system under analysis. Depending on the application, they may represent either the total number of links, as in the Erdös-Rényi model, or the degree sequence as in (Park and Newman, 2004; Garlaschelli and Loffredo, 2008; Squartini and Garlaschelli, 2011), or other topological properties (Squartini et al., 2013; Fronczak et al., 2013; Mastrandrea et al., 2014; Di Gangi et al., 2018; Becatti et al., 2019; de Jeude et al., 2018).

We rely on this literature in order to properly analyse ~10^6 tweets exchanged during the last Italian elections held on March 4, 2018. First we employ an undirected representation of the network of retweets by distinguishing certified from non-certified users. Then we identify groups of verified users by their interaction with the opposite layer, following the recipe of (Gualdi et al., 2016; Saracco et al., 2017): if two verified users are retweeted more than expected by the non-verified ones, they are likely to be related. Then, we analyse the community organisation of the resulting network and measure the polarisation of unverified users according to this division: as observed in other studies

(Dandekar et al., 2012; Flaxman et al., 2013; Bakshy et al., 2015; Nikolov et al., 2015; Quattrociocchi et al., 2014; Del Vicario et al., 2016b, 2016a; Schmidt et al., 2018a, 2018b), people tend to interact just with a single community, strongly polarising their opinions and we confirm this observation in the Italian elections of 2018. Finally, we study a bipartite and directed representation of the users' tweets and retweets network, in order to identify significant news consumers. In order to do so, we statistically validate the (directed) connections that cannot be simply explained by the "virality" of the tweets and the tweet/retweet activity of the users. This last validation is an extension of the approach of (Gualdi et al., 2016; Saracco et al. 2017) to direct bipartite networks.

Our analysis uncovers the various strategies for the electoral campaign, followed by different political alliances and highlights a different participation to the political debate, providing indications about the role of the in spreading viral content inside each community. Moreover, we observe not only the alliances presented *before* the election, but even a signal of the alliances *after* the elections. Indeed, our studies show the proximity of the electorate of the governing parties during the debate, despite belonging to different pre-elections alliances.

The rest of the paper is structured as follows: in Section 'Results' we describe the performed analyses with a report of the main findings; we summarise research questions, results and this work's contribution in the Section 'Discussion'. More detailed information regarding the dataset, the analyses and methods can be found in Section 'Methods' and in the Supplementary Material file.

Results

By means of the Python module `tweepy` we have used the Twitter Search API to download a sample of all tweets posted from January 28 to March 19, 2018. This tool has been adopted conscious of the fact that the scientific literature has largely investigated the problems related to sampling procedures in on-line social networks (Boyd and Crawford, 2012; Morstatter et al., 2013; Pfeffer et al., 2018), especially during political events as in (González-Bailón et al., 2014), and the difference with the alternative tool of the Streaming API, as in (Driscoll and Walker, 2014). The query has been performed only requiring each post to contain at least one of a set of Italian elections-related keywords, such as *elezioni*, *elezioni2018*, *4marzo*, *4marzo2018* (meaning elections, elections 2018, 4 march, 4 march 2018, respectively). It is worth noting that the API works in such a way that it returns a sample of the available tweets for each query, therefore the authors had no role in the selection of the tweets used for the analyses. For additional details regarding the nature of the dataset, refer to the Supplementary Material file, Section 'Introduction'.

The analysis conducted throughout the paper revolves around two main research questions: (1) Is it possible to infer the polarisation (i.e., the membership to a certain political wing) of an user from her/his relational activity instead of the textual information of her/his posts? (2) Once we detrend for the activity of users and the virality of the tweets, what can be said about the structure of the social network?

The former question is motivated by the intuition that social interactions of an user reveal her/his adherence to certain opinions. The second question is methodological: the analysis of a complex system is usually complicated by the presence of different types of noise covering the relevant information. Here, our purpose is to uncover the traits of the phenomenon that cannot be simply explained by the activity of the nodes in the system. A natural framework to address this challenge is an entropy-based

approach, which (by construction) provides an unbiased framework for the analysis of real networks (Squartini Tiziano and Garlaschelli Diego, 2017; Cimini et al., 2018).

In the following sections we address such questions. More specifically, Sections 'Identification of alliances via unverified user behaviours', 'Structure of the communities', 'Polarisation analysis' deal with the first question, while Section 'Influence analysis' investigates the second one.

Identification of alliances via unverified user behaviours. As a first step, we have split the sample of Italian-speaking users into two groups, the *verified* and *non-verified* users. This variable is directly available from the API and characterises those users that have requested to be authenticated by the system. Therefore the distinction between the two types is actually made by the users themselves, since in principle any account can request to be verified: when it does, Twitter guarantees that the account is authentic. For this reason, we expect this procedure to be mostly applied to those people considered of public interest, such as actors, politicians, newspapers, TV channels, radio channels etc.

After this distinction, we build the bipartite network of retweets between verified and non-verified users: an edge between two users indicates that one has retweeted the other's content at least once during the available time period. At this step we disregard edges' direction, therefore in principle we do not know who is the author of the post and who is the second one who shares the content. However, as shown in Fig. 4 of the Supplementary Material file, the actions of tweeting and sharing contents are mostly performed by verified accounts, while non-verified users mainly retweet already published posts.

In order to obtain groups of verified users based on their activity, we project the information contained in this bipartite network on the verified users layer. The result of the classical projection methods is a weighted monopartite network: two users are connected if they share at least one common neighbour on the opposite layer and the edge between them is weighted by the number of non-verified users who have simultaneously retweeted their posts. However, this method often generates a very dense projection; thus, several procedures have been proposed in order to establish the significance of the edges in the projected network, discounting different pieces of information. For this application we use as benchmark an entropy-based null model that discounts the information contained in the degree sequence of both layers (Saracco et al., 2017); in this way we are focusing on overlaps that cannot be explained by the activity of the users only. A brief description of this approach can be found in Section 'Methods', but refer to the Supplementary Material for the details of the entire procedure.

Given the projected and validated network of retweets, we have performed a reshuffled community detection procedure, i.e., the Louvain algorithm (Blondel et al., 2008) runs several times with a rearranged nodes' ordering and the partition with the highest modularity is selected: by doing so we overcome the original algorithm's order dependence (Fortunato, 2010). With this procedure we identify ten groups of non-isolated nodes. However, for the following analyses we only focus on a subset of four of them, being those with a remarkable number of nodes (more than a hundred) and a non-trivial interpretation. Two blocks identify quite well the groups of Movimento 5 Stelle (M5S hereafter) and right-leaning politicians. For instance, in the former we find the accounts M5S Camera, M5S Senato, M5S Europa and Movimento 5 Stelle, as well as the politicians Danilo Toninelli or Luigi Di Maio. Instead in the latter we see the accounts Forza Italia, Lega - Salvini Premier, Gruppo FI Camera, Fratelli d'Italia, Noi con Salvini, as well as the users Silvio Berlusconi, Matteo Salvini,

Renato Brunetta or Giorgia Meloni. The two remaining communities are instead more heterogeneous. In one of them, we find a high number of radios, newspapers or newscasts, such as Rai Radio 2, Radio 105, RTL 102.5, Tg Rai, Tg La7, Sky Tg 24, Rai News, la Repubblica, Il Corriere della Sera, Il Post. Since the majority of nodes refers to official accounts of news media, we characterise this group as the one collecting news spreaders and information channels. Finally, the last community encompasses some politicians within the left-leaning parties, such as Matteo Renzi and some other figures belonging to the Democratic party (included the account of the Partito Democratico itself). Therefore we use this interpretation for the last detected community.

Structure of the communities. Given the division in political alliances, we start analysing the topological characteristics of the subgraphs made by each group. In order to facilitate the understanding, we represent each community with a colour reminiscent of the colour used during the elections to represent the coalitions. We display the political coalitions (i.e., M5S, left-leaning and right-leaning alliances), respectively, in yellow, red and blue, while we use purple for the information channels community. This choice of colours will be used consistently throughout this manuscript.

In the present section we focus our analysis on some features of these identified clusters. Since the communities' average shortest-path-length are smaller or comparable to the values observed in the entire validated network, we suppose the presence of central users that behave as "bridges" between different communities in the process of news spreading. Table 1 provides a list of the first fifty most central users in the validated network of retweets, together with their affiliation with respect to the four identified groups. The first positions are covered by medias and journalists, the majority of which are affiliated with the purple community. However, we also observe users from the other communities, such as Il Fatto Quotidiano or Peter Gomez and IlSole24ORE.

Another important insight on users' behaviour comes from the observation of the hashtags used by the verified users. Excluding the set of keywords used to extract the data, Fig. 1 reproduces the hashtags used more frequently by the verified users of each community, selecting those with a frequency higher than or equal to the 0.5% of the total number of hashtags in the single community. All political (i.e., yellow, red and blue) communities have the name of their own party as the most mentioned tag: we observe, respectively, the hashtags "m5s", "pd" and "centrodestra" in their first position. Nevertheless, the second most used hashtag refers to the main opponent of the political alliance represented by the community. It is curious that this word is the Movimento 5 Stelle for both the right-leaning and left-leaning alliances, since it was effectively the most voted party at the elections. Instead, the second most used hashtag by Movimento 5 Stelle is "renzi", leader of the Partito Democratico at the moment of the elections, governing at that time. Also the major exponents of the other parties are mentioned, for instance "berlusconi", "salvini" and "dimaio" appear in the left-leaning and right-leaning parties, as well as in the M5S one and the mixed group. Even if the frequencies of the hashtags are comparable, all alliances have the name of their own coalition as the most frequent one, followed by the names of the major competitors. In three of the communities we also observe the tag "maratonamentana" as one of the most frequently used. This is the name of a TV show that is extremely popular during the elections, since it broadcasts live the scrutiny after the election day. This year it became even more viral, since the show was transmitted for almost 12 h without any interruption by the same anchor-man and started again after just 3 h of break. In the Supplementary Material file we graphically represent

another interesting behaviour regarding the use of tags, that will be better discussed in the following Sections.

Polarisation analysis. Once identified four groups of political actors and information channels representing the Italian political scenario, we proceed with the analysis of how the remaining accounts interact with them. In other words, the goal is to study whether the audience is polarised towards the source of information that better resembles their ideology or uniformly shares contents from accounts of different political orientation.

Following the notation introduced in the Supplementary Material file, we indicate with L and Γ, respectively, the sets of verified and unverified users collected in the two layers of the corresponding network, while we denote with $\mathcal{C}_c$ for $c = 1, 2, 3, 4$ the division in communities identified at the end of the phase above, denoting, respectively, the verified users belonging to M5S, information channels community, the left-leaning and right-leaning accounts. Also indicate with

$$\mathcal{N}_\alpha = \{i : i \in L \text{ and } m^*_{i\alpha} = 1\}$$

the set of neighbours of the non-verified user $\alpha \in \Gamma$ in the bipartite network of verified/unverified users, i.e., the set of verified users node α has interacted with. The polarisation index for α is

$$\rho_\alpha = \max(\{\mathcal{I}_{\alpha,c} : c = 1, 2, 3, 4\}) \text{ for } \alpha \in \Gamma \qquad (1)$$

with

$$\mathcal{I}_{\alpha,c} = \frac{|\mathcal{C}_c \cap \mathcal{N}_\alpha|}{|\mathcal{N}_\alpha|} \text{ for } c = 1, 2, 3, 4. \qquad (2)$$

The term $\mathcal{I}_{\alpha,c}$ denotes the fraction of α's interactions towards community c, i.e., the ratio of α's neighbours belonging to community c. This index has the following characteristics: is bounded in [0, 1], therefore $\rho_\alpha = 0$ means that no interaction has been observed with the four groups; values of ρ_α close to 1/4 indicate that user α equally interacts with the four clusters; for all the other values, the greater ρ_α the higher the inequality in the number of interactions with the four communities (i.e., ρ_α close to 1 means that user α almost always has interacted with the same group). The choice of this index has been mostly driven by the observation of users' interactions with the communities. See left panel of Fig. 2: each square in the heatmap reproduces the average value of the quantity on the x-axis, computed over the set of non-verified users belonging to the community on the y-axis.

This result indicates that most of the non-verified users have an extremely unbalanced distribution of their interactions with the members of the other alliances, since they mostly retweet content shared by people from their own community rather than from different ones. However, the definition of this polarisation index is independent on the number of interactions: two different users that, respectively, interact (for instance) one time or ten times towards the same community will be assigned the same value of polarisation. In order to take this situation into account, we show in Fig. 3 the distribution of the polarisation indices conditional on the total number of interactions observed for each user. In this case, the distributions are also concentrated towards higher values of the index. Moreover, for lower numbers of interactions, we also observe users that interact with people not belonging to the identified groups (this situation corresponds to the case of polarisation equal to zero). However, this behaviour disappears for higher numbers of interactions, meaning that more active users interact with the identified communities only, which translates into higher polarisation indices. Instead, the right-hand side of Fig. 2 represents the histogram of the polarisation values obtained for the non-verified users. As introduced by a

Table 1 First most central verified users: list of the most central users in the validated network of verified users

community	users	centrality index
●	la Repubblica	0.0532
●	Agenzia ANSA	0.0382
●	Il Fatto Quotidiano	0.0300
●	Sky TG24	0.0292
●	Corriere della Sera	0.0285
●	YouTrend	0.0247
●	Rainews	0.0228
●	Peter Gomez	0.0152
●	Il Post	0.0142
●	IlSole24ORE	0.0137
●	Matteo Renzi	0.0133
●	Tg La7	0.0124
●	Ferruccio de Bortoli	0.0117
●	Marco Castelnuovo	0.0104
●	Internazionale	0.0100
●	Il Foglio	0.0100
●	Lorenzo Pregliasco	0.0095
●	Confindustria	0.0093
●	Agi Agenzia Italia	0.0086
●	Vittorio Zucconi	0.0085
●	Selvaggia Lucarelli	0.0072
●	Pietro Grasso	0.0071
●	Tgcom24	0.0069
●	Partito Democratico	0.0067
●	diMartedì	0.0066
●	La Stampa	0.0066
●	Stefano Feltri	0.0066
●	L'Espresso	0.0065
●	Giuseppe Smorto	0.0063
●	Ficarra e Picone	0.0062
●	Rai Radio1	0.0061
●	Il Viminale	0.0061
●	Ezio Mauro	0.0057
●	Movimento 5 Stelle	0.0057
●	Anna Ascani	0.0055
●	insopportabile	0.0055
●	Agenzia DIRE	0.0054
●	Matteo Salvini	0.0053
●	Fabio Chiusi	0.0052
●	Riccardo Puglisi	0.0052
●	Luca Tentoni	0.0051
●	Paolo Attivissimo	0.0045
●	RTL 102.5	0.0045
●	Franz Russo	0.0044
●	Matteo Pedrosi	0.0042
●	Marco Travaglio	0.0041
●	trashitaliano.it	0.0041
●	Tommaso Labate	0.0040
●	Giorgio La Porta	0.0040
●	Sebastiano Messina	0.0038

The coloured dot left to the name indicates the political affiliation of the user according to our division in community
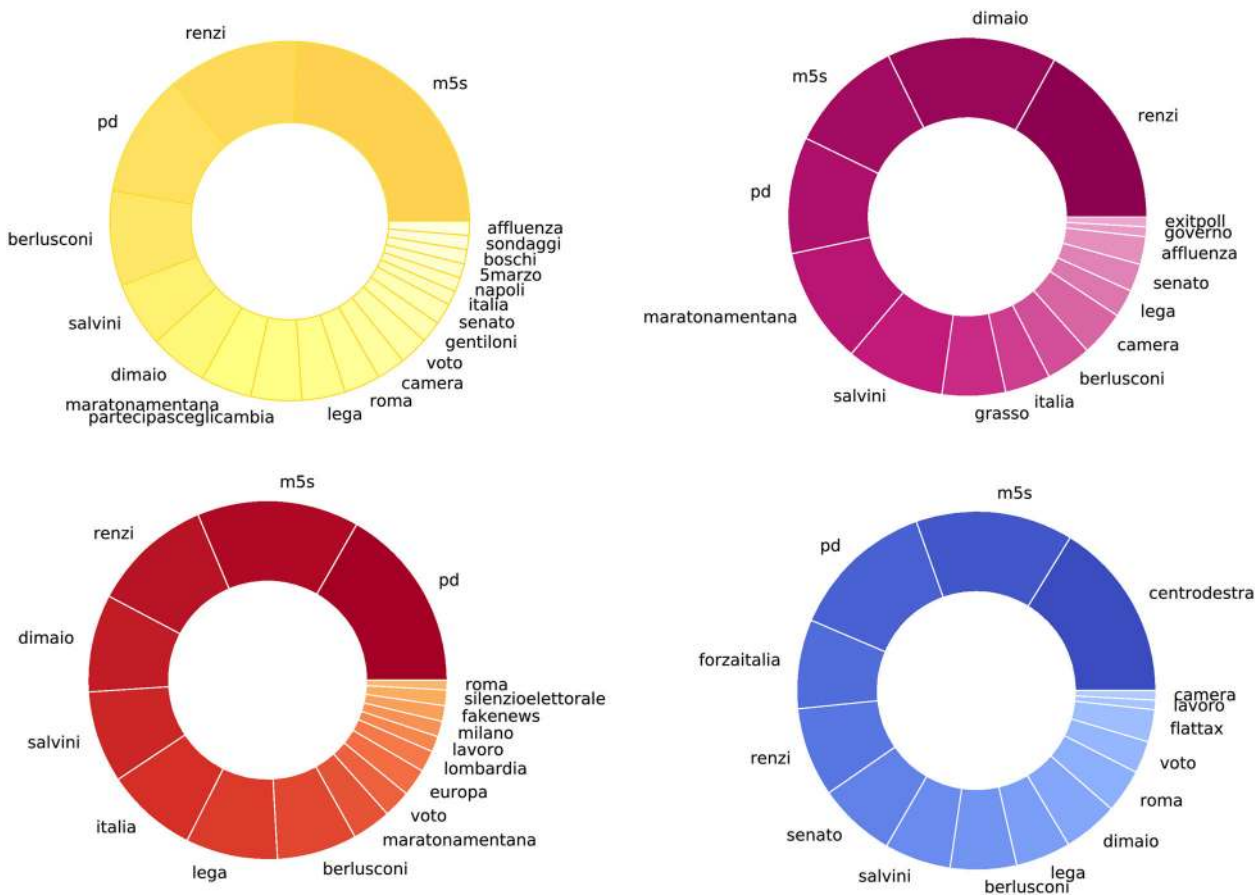
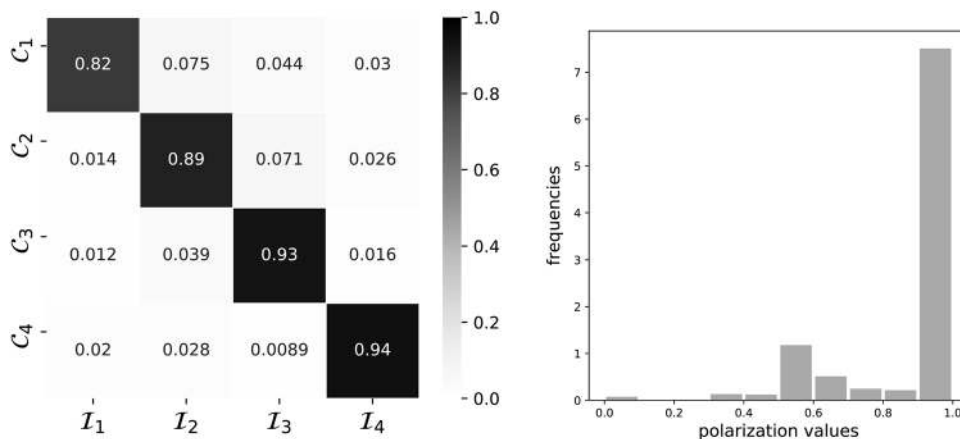**Fig. 1** Hashtags used by the four network communities

Fig. 2 Users' interactions

preliminary evaluation of users' behaviour, many non-verified users show high polarisation values, meaning that their attention patterns are mostly focused towards a limited group of political actors.

Figure 4 shows the biadjacency matrix of the bipartite network of verified and non-verified users. The coloured blocks identify the four communities obtained with the community detection method: the red and blue blocks, respectively, identify the groups of left-leaning and right-leaning politicians; the yellow community collects the available political figures within the M5S party, while the violet group represents the information channels community. The rows of the matrix have been ordered according

to the division in communities of the verified users, while the non-verified users have been sorted according to their political affiliation, i.e., the number of interactions towards each group, that is equivalent to the computation of the numerator of the term in Eq. (2) for all the available communities. Such a matrix exhibits a block structure along the diagonal, indicating a greater number of interactions towards the "preferred" community with respect to the others and therefore a higher density of links within the blocks with respect to the external density.

Once the non-verified users have been classified in political alliances based on their polarisation values, we perform again an analysis of the most used hashtags in each group. The result is
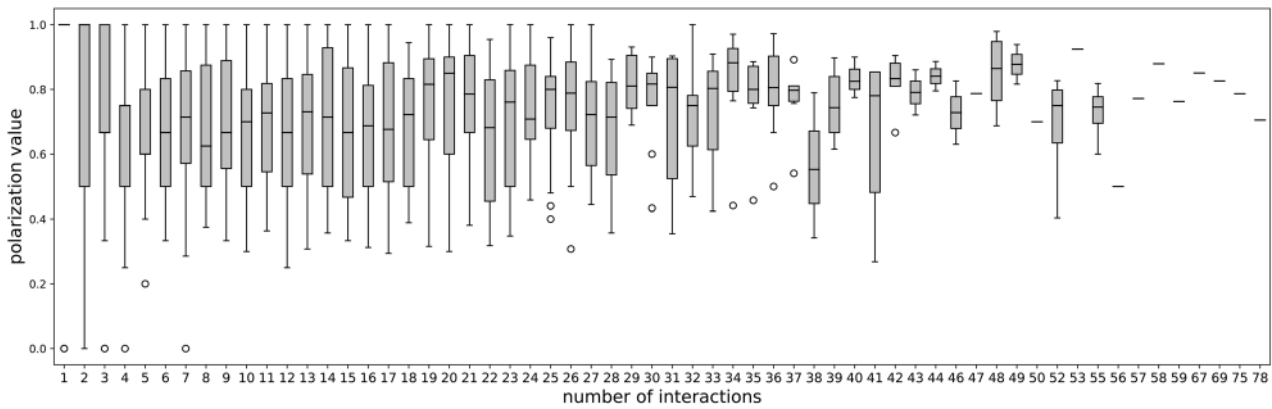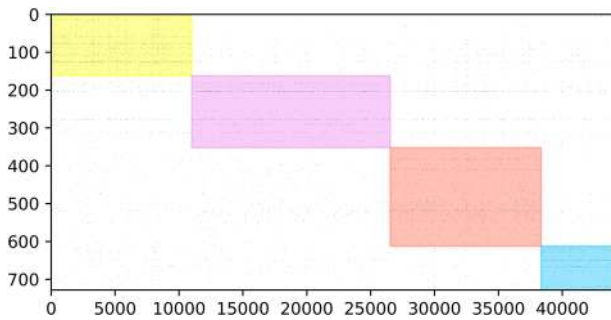
Fig. 3 Users' polarisation

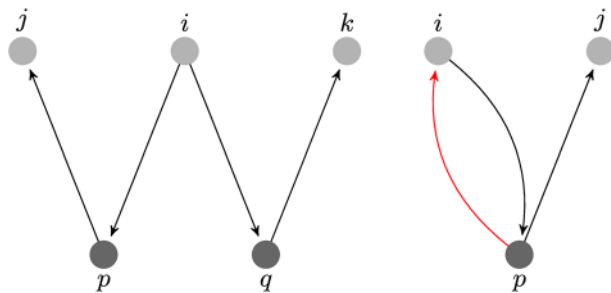**Fig. 4** Biadjacency matrix showing the communities of users

Fig. 5 Bipartite and directed representation of the network of retweets

shown in Fig. 1 of the Supplementary Material file. In this case the result is interesting: the analysis of the hashtags used more frequently by *all users* in each coalition (verified and non-verified) reveals that the tag with the highest frequency is related to the opponent group, as if the membership to a certain community could also be driven by the dislike of the opponent group rather than by the approval of the proposal of one's network-based alliance only.

At this point, we think it is worth performing a more in-depth analysis of the concept of polarisation. Broadly speaking, the concept of *polarisation* in social sciences refers to a division of the society under analysis in several divided groups, characterised by different beliefs or opinion. Instead, the formation of *echo chambers* can be seen as a consequence of a polarised society: because of the fact that people only interact with other members of their same social group, news and beliefs are reinforced and amplified since they reverberate within the same closed social sphere. Therefore, we consider the results of our analysis more related to the phenomenon of polarisation: given our definition, a

high value of this index indicates that an user systematically interacts with (i.e., retweets content shared by) people belonging to her/his same coalition rather than with users from different alliances, but we disregard any information concerning the content of the shared news and we do not take into account how it is amplified by being shared within the same group of people.

Finally, we think it is worth noting that, contrarily to the majority of studies on the same subject, the polarisation of non-certified users has not been studied simply labelling the units in the verified accounts layer. On the contrary, users in the same community share a significantly high number of non-verified users that have retweeted their content. Indeed, a validated edge between users i and j indicates that a high number of non-verified accounts who shared contents posted by i has also retweeted posts published by j and therefore they are considered similar by a majority of their audience and followers. Therefore, in our analysis, the driver to understand the division in clusters is solely the behaviour of non-certified accounts.

Influence analysis. At this point, we proceed with the identification of the significant sources of Twitter viral content. Inspired by the work of (Bovet and Makse, 2018), we propose a *bipartite and directed network of information flow*. An example of this kind of graph is provided in Fig. 5. The users on the upper layer tweet and retweet the posts represented on the lower layer. An outgoing edge (i, p) as in the figure indicates that user i has published tweet p during the considered time period, while a link in the opposite direction (p, j) denotes that user j has retweeted the same post p. In this situation, j behaves as a *spreader*: the larger the number of spreaders of an user, the larger the audience of the contents shared by her/him. Moreover, the larger the number of tweets posted by i that have been retweeted by j, the tighter the social bond between the two.

As in (Bovet and Makse, 2018), we simply project the bipartite and directed network of tweets and retweets onto the users layer: a directed edge between i and j in the projected graph indicates that j has retweeted i's posts at least one time. A list of the nodes with the highest out-degrees per community can be found in Table 2. The ✓ next to the name indicates verified users. Clearly the top 20 most-retweeted users in each group are mainly certified, because they are extremely active with the purpose of enlarging their pool of voters. From the same graph we also select the top 20 nodes with the highest in-degrees, i.e., the spreaders, who have a high number of retweets. In this case, the most retweeting users are mostly uncertified. Moreover, we also detect some of the involved accounts that have been suspended, suggesting the presence of accounts violating the Twitter policy

Table 2 Most retweeted users: Following the approach of (Bovet and Makse, 2018), we build the monopartite and directed network of retweets, in which an edge (*i*, *j*) indicates that *j* has retweeted *i* at least once

Community 🟡	Community 🟣
Il Fatto Quotidiano ✓	la Repubblica ✓
Peter Gomez ✓	Agenzia ANSA ✓
NVPD	trashitaliano.it ✓
EL	Sky TG24 ✓
Manlio Di Stefano ✓	YouTrend ✓
KS	Corriere della Sera ✓
Marco Travaglio ✓	Il Post ✓
AB	Rainews ✓
Ficarra e Picone ✓	Tg La7 ✓
Movimento 5 Stelle ✓	Pietro Grasso ✓
Du	Rolling Stone Italia ✓
Da	Repubblica Tv
SiVV	Sh
LC	Pa
AK	Fran Altomare ✓
Selvaggia Lucarelli ✓	HaP
EC	Tg2
Ferruccio de Bortoli ✓	Rai Radio2 ✓
SeVV	Alessandro Masala ✓
carlo sibilia ✓	Giuseppe Civati ✓
LM	DoC
CL	Alessandro Alciato ✓
Nicola Morra ✓	Daniela Collu ✓
Community 🔴	Community 🔵
Matteo Renzi ✓	Matteo Salvini ✓
Partito Democratico ✓	Claudio Borghi A. ✓
Vittorio Zucconi ✓	Giorgio La Porta ✓
CC	Giorgia Meloni ✓
EC	Silvio Berlusconi ✓
IlSole24ORE ✓	Dr. Alice Weidel ✓
Sebastiano Messina ✓	PR
Alessia Morani ✓	Forza Italia ✓
Il Foglio ✓	Lega - Salvini Premier ✓
Pietro Raffa ✓	CasaPound Italia
JI	BR
Riccardo Magi ✓	SDS
DaC	Renato Brunetta ✓
mattia feltri ✓	Al
Lorenzo Pregliasco ✓	AM
Anna Ascani ✓	SG
CZ	Fratelli d'Italia ✓
Internazionale ✓	Tgcom24 ✓
Carlo Calenda ✓	Gruppo FI Camera ✓
Marco Castelnuovo ✓	AC
Davide Serra ✓	ForzaItalia ELEZIONI
Marco Cappato ✓	LB
insopportabile ✓	AR

Then we select the directed subgraph generated by each community and here we report the list of the most retweeted users within each subgraph. The ✓ next to the name indicates that the account has been verified
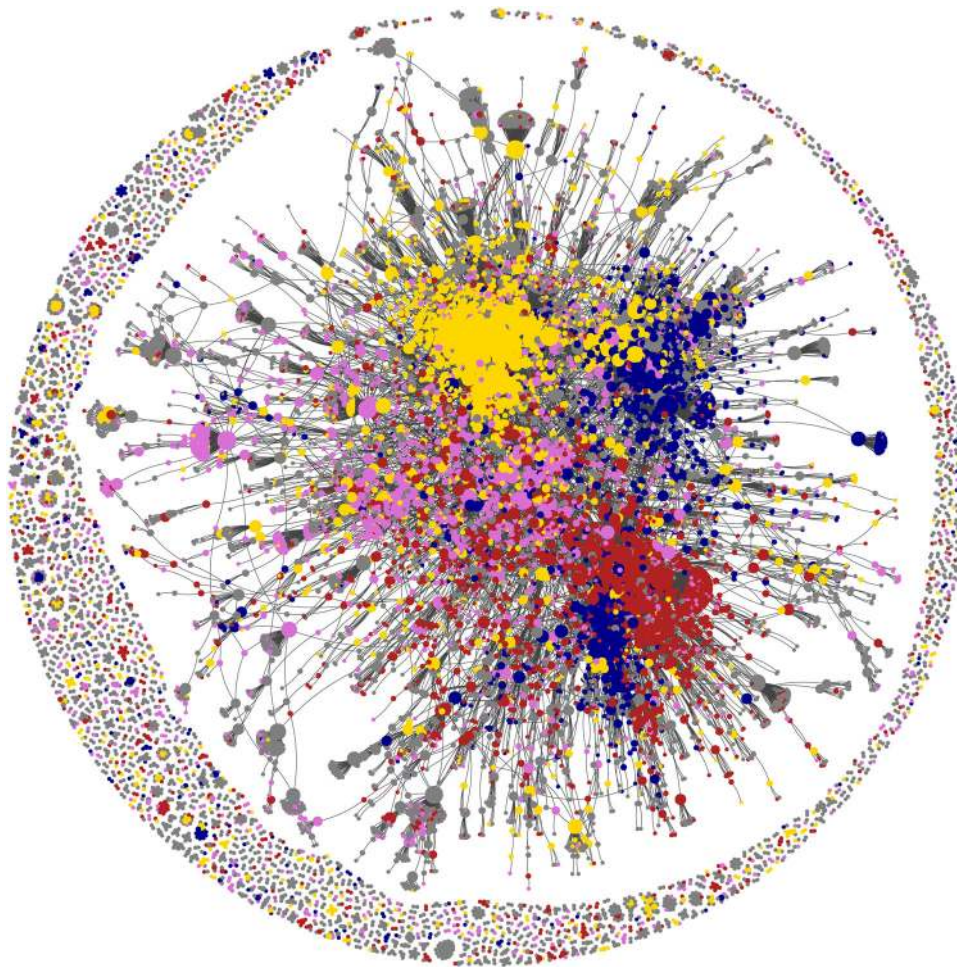
Fig. 6 Validated and directed network of retweets

during the event of the elections: there is one suspended account in the M5S coalition and two others in the right-leaning one.

However, these findings are not really explanatory in identifying the most *effective* users in the directed network of retweets, since we are missing a benchmark for stating if j is a *significant* contributor of the popularity of i's posts. The identification of such significant tights will be performed in two steps: first, we define a suitable benchmark model to evaluate the significance of pairwise connections; this can be done by using the Bipartite Directed Configuration Model proposed in the Appendix of (de Jeude et al., 2018), thus discounting the information related to the in- and out-degrees of nodes belonging to both layers (in the present case the information of in- and out-degrees of posts and users). Then we extend the validation presented in (Saracco et al., 2017) to this kind of graphs, in order to identify the significant tights between source and spreader and validate the directed network of information flow. The entire procedure is explained in Section 'Methods', but additional details of the method are presented in the Supplementary Material file. In the final validated and directed network of users, only the pairs of users in which one of the two retweets the other's contents more than what is expected by the considered null model will be connected. A pictorial representation of the network of validated retweets is provided in Fig. 6.

Nodes' colour identifies the user's community while nodes' dimension is proportional to their out-degree in the validated graph, i.e., the number of significant spreaders they have. The

structure of this network is better represented in Figs 7–10. Each plot focuses on the structure of the subgraphs of the directed network generated by each community. Nodes' dimension is again directly proportional to their out-degrees in the subgraph: the larger the node, the higher the number of times that user has been retweeted by the other accounts. Nodes' colour is instead related to whether the account has been verified or not: blue for verified users, orange for non-verified ones.

The first plot (Fig. 7) is related to the M5S community and shows a strongly connected block of (mostly non-verified) nodes that retweet among themselves and with the verified accounts of the community, the most central of which are the Twitter accounts of the newspaper Il Fatto Quotidiano and its journalists Marco Travaglio, Peter Gomez and Antonio Padellaro. In the other communities journalists and newspapers do not form such a strong core. It is interesting to note that the M5S political leader Luigi Di Maio does not belong to this central community but is located in a small community outside this large component.

The second plot represents the purple community (see Fig. 8). The most central nodes are the verified accounts of newspapers (see for example La Repubblica or Il Corriere della Sera) and information channels (such as Sky TG24, Tg La7, Agenzia Ansa or Rainews). Some politicians such as Pietro Grasso or Giuseppe Civati are present in this group, together with the political parties of Rifondazione and an extreme-left party (which account is not verified). These politicians represent the most extreme, left-leaning orientation who have not encountered a commonality of
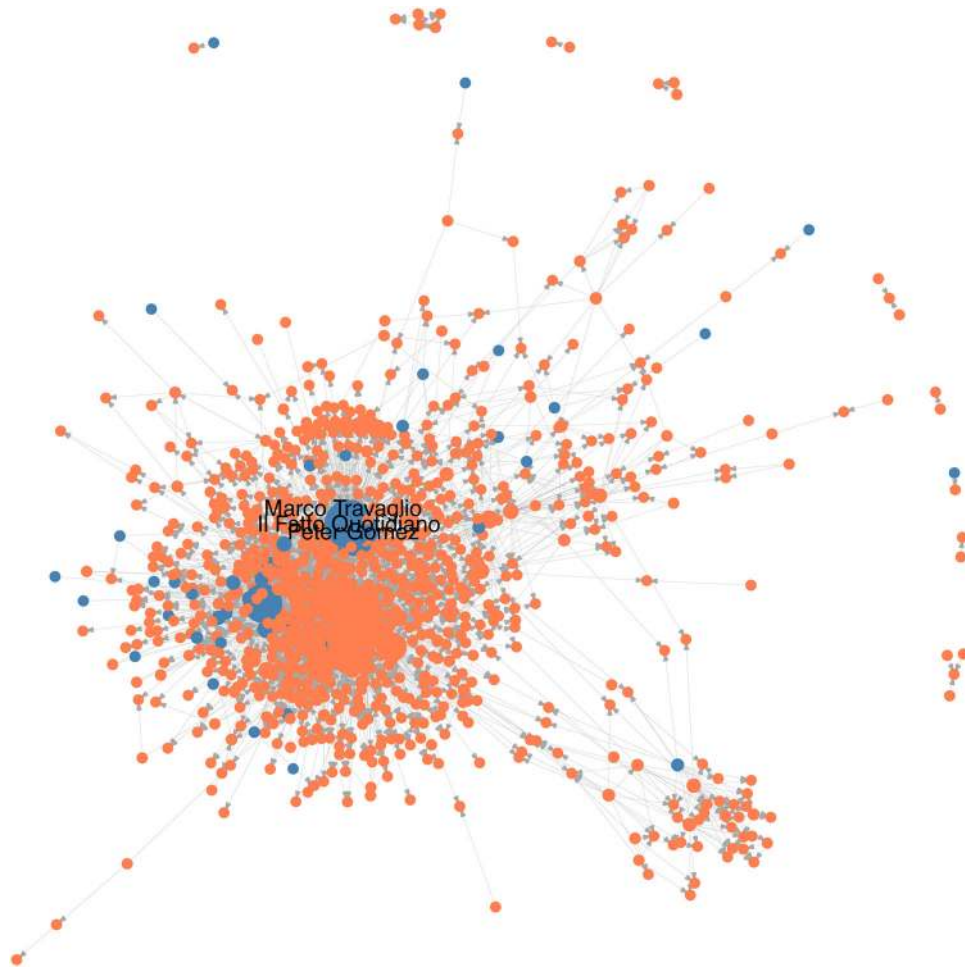
Fig. 7 Directed subgraph for community $\mathcal{C}_1$

interests and supporters with the accounts in the community of Partito Democratico and therefore they belong to different communities.

In the plot associated to the left-leaning community (see Fig. 9) we identify a central block of mostly non-verified users. The most retweeted figures are Matteo Renzi and the account of Partito Democratico, as shown by their high values of out-degrees. The remaining verified nodes are mostly well-known characters in the political scenario (such as Maria Elena Boschi and Carlo Calenda), as well as newspapers (see for example Il Foglio or IlSole24Ore). Among the non-verified users we have the accounts of the Partito Democratico political parties related to the areas of Milan and Rome.

Finally, the plot associated to the right-leaning community (see Fig. 10) is characterised by two quite separate clusters; one of them is centred on the accounts of people belonging to Lega Nord, such as Matteo Salvini, Claudio Borghi and the party Lega-Salvini Premier. On the other side there are the accounts of Forza Italia and Gruppo FI Camera and some of its exponents like Silvio Berlusconi or Renato Brunetta. The two verified nodes of Giorgia Meloni and Fratelli d'Italia (the political party she is leading) receive retweets from both sides, nevertheless being closer to the Lega pole. Another popular node is a neo-nazi party (which account is not verified), that has its own circle of retweeters and share some interactions with the subgroup of Lega Nord.

In order to understand how viral news propagate through the validated network, we analyse the percentage of retweets coming from inside of the same community and the percentage that,

instead, derives from the other ones. The results are shown in Fig. 2 of the Supplementary Materials and are extensively commented therein.

Methods
Definition of the polarisation network
Bipartite network of verified/unverified users. As a first step, we have split the sample of available and Italian-speaking users in two categories, the groups of *verified* and *non-verified* users. Each account can request to be verified by the system: by doing so, Twitter guarantees that the account is authentic. This kind of procedure is, in general, applied to all those people who are considered of public interest. Therefore, we expect the accounts of famous people, politicians, newspapers, TV channels, radio channels etc. to be included into the set of verified users, while all the remaining users to belong to the other set.

We do not have specific information regarding the type of accounts that use Twitter daily and which of them may ask to be verified. However, we know from (AGCOM, 2018) that social networks are increasingly used in Italy with the purpose of retrieving information. In particular, the report shows that there is a difference in the type of social networks used by citizens against the ones consulted by professionals: on-line social networks such as Facebook, Instagram, Snapchat are used by all people, professionals, as well as common users, while others such as Twitter or LinkedIn are more popular among business people (the report refers to journalists). Indeed, journalists retrieve
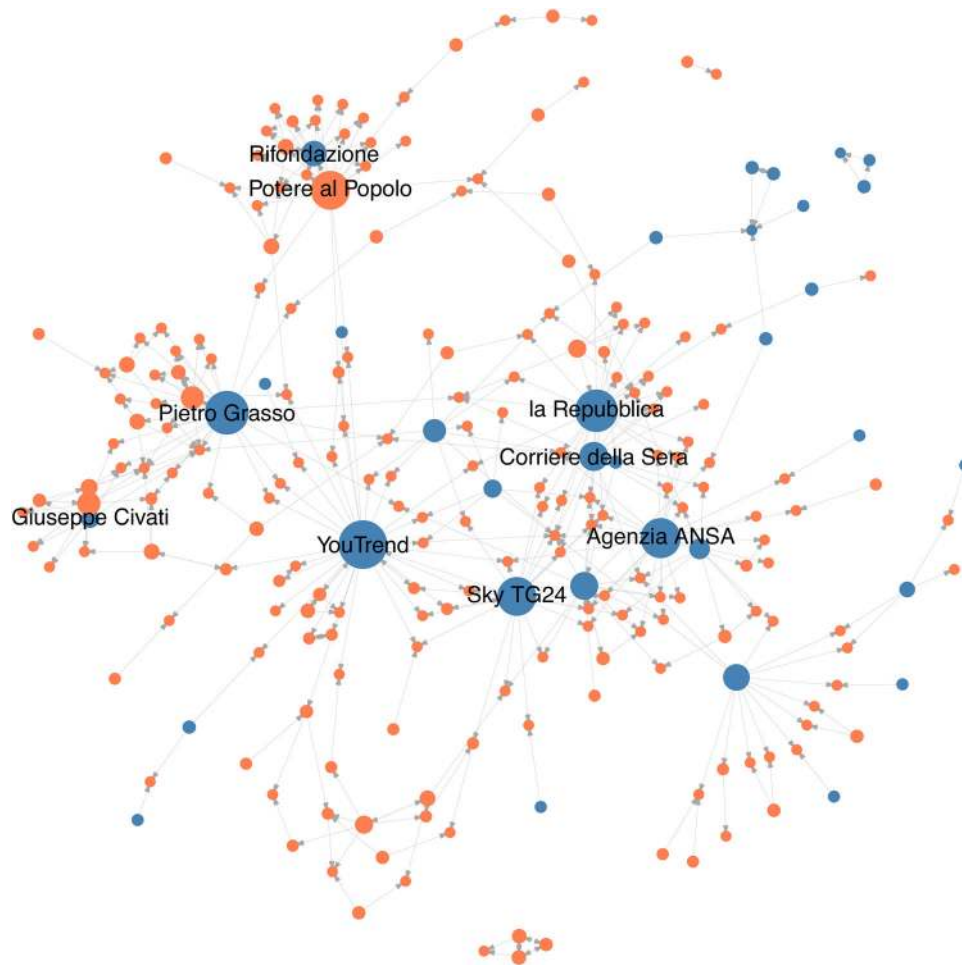
Fig. 8 Directed subgraph for community $\mathcal{C}_2$

information on Twitter, share their own and fellows' articles and it is also a way to keep an eye on users' discussion on a specific topic. Therefore, we may expect to find professionals such as politicians, newspapers, information channels like TV or radios to belong to the set of verified accounts.

Given this division, we construct a first bipartite network from the data, the network of retweets between verified and non-verified users during the whole period, denoted as $\mathbf{G}_{Bi}^* = (L, \Gamma, E)^2$. By definition of bipartite networks, with this representation we exclude all the cases in which two certified (or non-certified) users retweet each other. We do so since we are interested in exploiting the way normal people consume the news to detect connected groups of certified users. Refer to Section 4.1.1 of the Supplementary Materials for a more detailed description of this network.

Bipartite configuration model and undirected validated projection. In order to evaluate the relevance of the bipartite network of retweets between verified and non-verified users, we need to define a suitable model as a benchmark for our measurements. Indeed, at the present step we want to discount the information due to the activity of the users, both verified and unverified, in order to detect non-trivial superposition of connections. Thus, we implement a Bipartite Configuration Model (BiCM) (Saracco et al., 2015) for this case, that maximises the entropy of the system, constraining the degree sequence of the two layers. A full detailed description of the null model can be found in the Supplementary Materials. Let us just remind here that at the end of this procedure we obtain a probability per graph that factorises in independent probabilities per link.

In order to infer information about similar behaviours of verified users, we first project the information contained in the bipartite network $\mathbf{G}_{Bi}^*$ on the verified-users layer. The result of the classical projection methods is a weighted monopartite network: two users $i, j \in L$ are connected if they share at least one common neighbour on the opposite layer Γ and the edge between them is weighted by the number of non-verified users who have simultaneously retweeted their posts. This quantity is expressed by the number of V–motifs between users i and j (Diestel, 2012; Saracco et al., 2015)

$$V_{ij}^* = \sum_{\alpha \in \Gamma} m_{i\alpha}^* m_{j\alpha}^*.$$

However, this method often generates a highly connected projection; in order to extract the statistically significant information of this projection we implement the method of (Saracco et al., 2017). Roughly speaking, the method consists in comparing the realised V–motifs with the expected value of the Bipartite Configuration Model. In this sense, the statically significant V–motifs are extracted: otherwise stated, all motifs that cannot be explained by the degree sequence only are validated. More details can be found in the Supplementary Materials, Sections 4.1.2–4.1.3.

Directed network of information flow
Bipartite directed network of tweeting/retweeting activity. In this section, we describe the construction of the bipartite and directed network of information flow. A directed edge (i, p) indicates that user i is the original creator of post p, while an edge on the
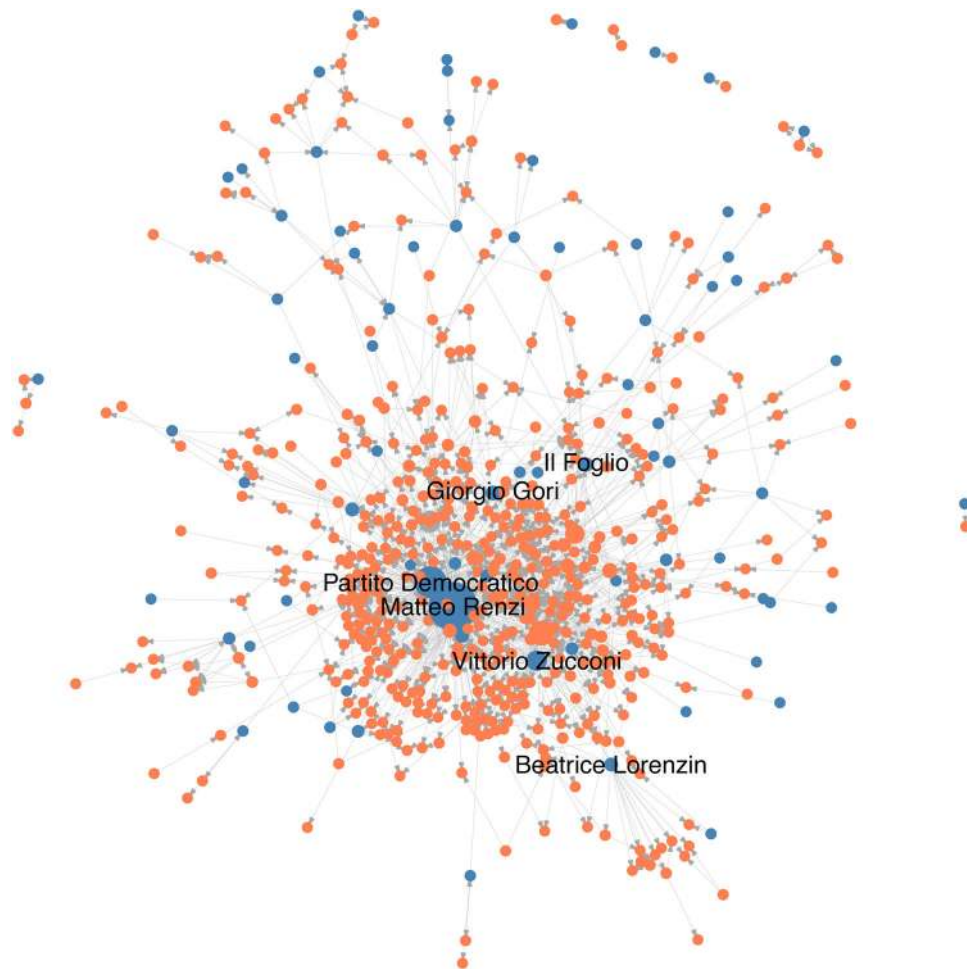
Fig. 9 Directed subgraph for community $\mathcal{C}_3$

opposite direction (p, j) shows that j has retweeted tweet p *at least one time* during the elections days. Such a situation is shown in the left panel of Fig. 5. In the picture, i is an user who publishes two posts p and q during the election days; user j retweets post p and user k retweets post q. Notice that, by construction, we cannot keep track of chains of sequential retweets (consult Section 1 of the Supplementary Materials for additional details about the nature of the data set). Therefore, the edge (q, k) does not necessarily mean that k has retweeted directly from i's post: it is possible (though less likely) that k has retweeted post q from one of user i's retweets. Another particular case is self-retweet: Twitter allows users to retweet their own posts, either directly from the original tweet or from somebody else's retweet. With this network representation, these cases can be illustrated as in right panel of Fig. 5, where the retweet by i itself is indicated with the red arrow. However, these type of edges have been excluded from the analysis, since they represent a very small percentage of the overall number of retweets observed in the data and we are rather interested in how people from the same political coalition interact with each other to boost the visibility of their opinion.

This bipartite and directed network $\mathbf{G}^*_{\text{BiD}}$ can be represented by means of two biadjacency matrices, one for the tweets $\mathbf{T}^* = \{t^*_{ip} : i \in U \text{ and } p \in P\}$ and the other one for the retweets $\mathbf{R}^* = \{r^*_{pj} : p \in P \text{ and } j \in U\}$. Essentially $t^*_{ip} = 1$ if user i has tweeted post p and 0 otherwise and $r^*_{pj} = 1$ whether at least one retweet to post p by user j is observed during the observation period. Refer to Section 4.2 for a more detailed description of this directed network.

Bipartite directed configuration model and directed validated projection. For the present analysis we are interested in identifying the most pervasive tweet flows, discounting both the users' activity and the tweets' virality. Otherwise stated, the goal is to measure which are the non-trivial patterns observed in the network, not explained by the degree sequence of both layers, to highlight and describe the structure of the system. The randomisation procedure for a bipartite directed network, imposing constraints on the in-degree and out-degree sequences, was first presented in the Appendix of (de Jeude et al., 2018); we revise the procedure that leads to the definition of the probabilities in the Supplementary Materials. Let us remark that, as in the previous (undirected) case of the BiCM, the probability per graph factorises in terms of probabilities per (directed) link. Moreover, with respect to the previous case, there is a crucial simplification: each post cannot have more than an author, thus the in-degree of tweets over the ensemble defined by Bipartite Directed Configuration Model (BiDCM) is simply $\langle \kappa^{\text{in}}_p \rangle = 1$.

Extending the definition of V–motifs to directed networks, the quantity $\mathcal{V}_{ij}$ identifies the number of times j has acted as a spreader for i, retweeting content posted by her/him[3]. Using the null model's expected value of this quantity we are able to infer the statistical significance of the tight between i and j, following a procedure similar to the one of (Saracco et al., 2017); consult the Supplementary Materials for further details. This procedure returns a squared $N_U \times N_U$ matrix collecting the significance of each link and each edge is treated as a separate null hypothesis to test[4]. Only the links associated to rejected hypotheses have to be
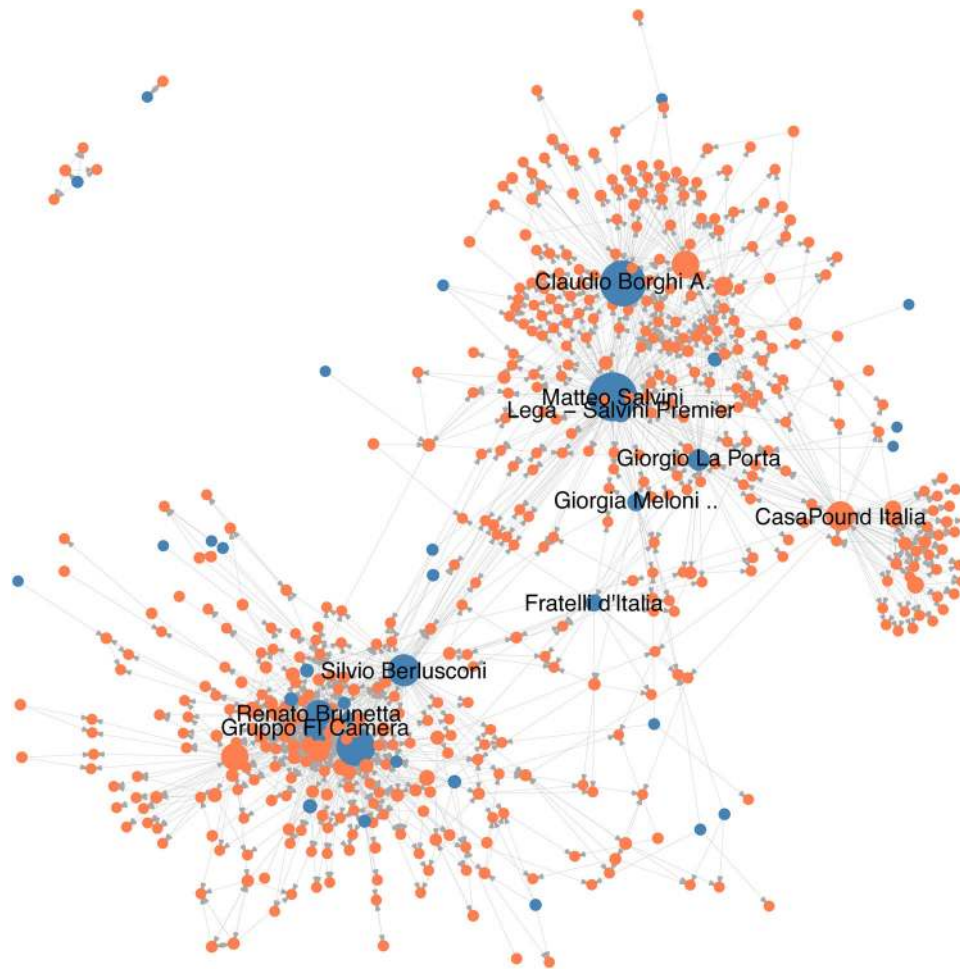
Fig. 10 Directed subgraph for community $\mathcal{C}_4$

included in the validated network: whenever $\mathcal{V}_{ij}^*$ is statistically validated, j is considered a significant spreader of i's tweets and the directed edge (i, j) is included in the validated network of information flows $\mathbf{F} = (f_{ij})_{ij \in U}$. Refer to Sections 4.3–4.4 for a more detailed description of these methods.

Discussion

In this work, we focused our attention on two main research goals: the detection of potential political alliances starting from the observation of the on-line behaviour and the identification of the sources of Twitter viral content from the tweeting and retweeting activity. We have addressed these tasks on the on-line social network built during the last Italian electoral campaign of 2018, gathering the data from the Twitter API during the month before the elections (held on March 4, 2018).

As a first step, we exploited the way people consume news on the social media to identify four groups of strongly connected verified users: two certified users are connected in the validated network of retweets if a significantly high number of non-certified accounts retweets content published by both of them. By construction, the behaviour of non-certified users is exploited to understand the division of the political sphere into clusters. An analysis of the obtained communities shows that the most central accounts are mostly newspapers, journalists and information channels in general, each of them belonging to one of the previously listed clusters. Looking at the hashtags included in the posts published by the verified users in each group, we see that

the most used keywords are referred to the party itself and its members, followed by keywords related to political competitors. However, when also non-verified users are included in the analysis, the behaviour changes and the most popular keyword becomes the name of the major competitor. Given the obtained division in political alliances, we studied the behaviour of the remaining non-verified users towards these groups. More specifically, we have observed the fraction of retweets directed towards each alliance: we observe a strongly polarised behaviour, since the majority of the uncertified accounts in the bipartite network of retweets mostly interact with one community only. In order to strengthen this result, we also perform a different analysis comparing the distribution of polarisation values observed for users with the same number of interactions. Also in this case, the distribution is skewed towards the higher numbers, indicating a focus towards the same group of users.

As a second step, we focused our attention on the identification of significant news spreaders. Following the methodology presented in (Bovet and Makse, 2018), we constructed the directed network of retweets among users and selected the names with the highest out-degree and in-degree. In order to statistically validate our findings, we constructed the bipartite and directed network of tweets and retweets introduced in Sections 'Bipartite directed network of tweeting/retweeting activity', 'Bipartite Directed Configuration Model and directed validated projection' and performed the validation procedure described therein. The outcome of this analysis is a monopartite, directed network of users, in which an edge from i to j indicates that the latter retweeted contents posted by the former a

significantly high number of times. The visualisation of this vali-
dated network helps to understand the actual composition of each
coalition, as well as the possible interconnections between them.
For instance, we observed that part of the connections between
one community and another one happens between verified and
unverified users, where, in most of the cases, the latter retweeted
some posts from the former. However, we also see connections
involving newspapers and information channels belonging to
different coalitions, confirming again their essential role and
centrality in spreading news on the social networks. Finally, we
analysed the origin of the retweets received by each community:
even though the distribution seems less polarised in this case, a
higher percentage of the retweets received by a community comes
from users belonging to the community itself, especially for the
cases of $\mathcal{C}_3$ and $\mathcal{C}_4$, showing that most of the interactions still
comes from the same sphere of influence (see Figs 2 and 3 of the
Supplementary Material).

In our view, the methodological contributions of this paper are
manifold. First, at odds with a majority of papers dealing with the
same topic, our dataset was not manually labelled: we identified
groups of strongly connected users starting from the behaviour of
non-certified ones and their interactions with certified people.
Despite this data-driven approach, we manage to identify four
clusters of users that are closely aligned with the known Italian
political division. Therefore, how people consume the news and
interact with the main political figures helps to shed light on the
actual division of the verified users according to their political
orientation. Our second contribution resides in the representation
of the network of activities on Twitter as a bipartite, directed
network. We employed the null model for directed bipartite
networks proposed in the Appendix of (de Jeude et al., 2018) to
identify the significant information flows between pairs of users
and different communities. We observe that the right-leaning
alliance is divided in two subcommunities: one centred on Ber-
lusconi and Forza Italia, closer to the Partito Democratico com-
munity; the other, led by Salvini and Lega party, closer to the M5S
community. The result is striking since M5S and Lega actually
allied to form a new government, after the elections, when no
predefined alliance obtained the absolute majority. Again, a
reminder might be worth concerning the fact that our results are
obtained from the analysis of a sample of tweets. Therefore, given
the issues related to Twitter Search API, the outcomes of the
analysis might be affected by the employed sampling procedure.

An important problem in the analysis of complex systems is the
detection of relevant signals with respect to the random noise, that
typically affects the interactions among a huge amount of different
agents. In this sense, in order to filter such a contribution from the
real system, several approaches have been proposed. The variety of
these proposals resides in the nature of the random noise and in
how such component should be discounted. In our work, this kind
of analysis allows to see that signals of the political alliances estab-
lished after the elections were already present in the on-line sphere
before the elections and could be detected when observing the way
people consume the news on Twitter. Interesting lines of research,
left for future investigations, include a validation of our approach to
a corpus of different elections, in order to identify potential reg-
ularities or differences between countries, and the use of tools from
natural language processing to infer how positively connoted and
negatively connoted tweets are distributed across the communities.

Data availability
The datasets generated and analysed during the current study are
not publicly available due to the data cleaning procedure the
authors implement for research purposes, but are available from
the corresponding author on reasonable request.

Notes
1 Nodes play the role of the volume in the canonical definition in statistical physics.
2 In the following we will refer to quantities regarding the real network with an asterisk *.
3 Note that $\mathcal{V}_{ij} \neq \mathcal{V}_{ji}$.
4 Actually, there is a tricky part. Since all users that were not author of retweeted posts have
probability exactly equal to zero to spread a tweet, their probability to be the source of a
directed $V-$ motif is again exactly zero, for each of the possible tweet. Those cases were
not considered in the FDR.

References
Adamic LA, Glance N (2005) The political blogosphere and the 2004 U.S. election.
In: Proc. 3rd International Work Link Discovery-LinkKDD '05. ACM Press,
New York, p 36–43
AGCOM (2018) Report on the news consumption. Technical Report. https://www.
agcom.it/documents/10179/10214149/Studio-Ricerca+10-04-2018/
4619854b-6d9b-491b-a3c8-f7444487006d?version=1.0
Badawy A, Ferrara E, Lerman K (2018) Analyzing the Digital Traces of Political
Manipulation: The 2016 Russian Interference Twitter Campaign. https://doi.
org/10.1145/nnnnnnn.nnnnnnn. http://arxiv.org/abs/1802.04291
Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news
and opinion on Facebook. Science 348(6239):1130–1132. https://doi.org/
10.1126/science.aaa1160
Barabási A-L (2011) The network takeover. Nat Phys 8:14–16. https://doi.org/
10.1038/nphys2188
Becatti C, Caldarelli G, Saracco F (2019) Entropy-based randomization of rating
networks. Phys Rev E 99(2):022306
Bekafigo MA, McBride A (2013) Who Tweets About Politics?: Political Partici-
pation of Twitter Users During the 2011Gubernatorial Elections. Soc Sci
Comput Rev 31(5):625–643. https://doi.org/10.1177/0894439313490405
Berelson BR, Lazarsfeld PF, McPhee WN (1954) Voting: a study of opinion for-
mation in a presidential campaign, University of Chicago Press, Chicago
Bindi J, Colombi D, Iannelli F, Politi N, Sugarelli M, Tavarone R, Ubaldi E (2018)
Political Discussion and Leanings on Twitter: the 2016 Italian Constitutional
Referendum. http://arxiv.org/abs/1805.07388
Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of
communities in large networks. J. Stat. Mech. Theory Exp. 2008(10). https://
doi.org/10.1088/1742-5468/2008/10/P10008
Borondo J, Morales AJ, Losada JC, Benito RM (2012) Characterizing and modeling
an electoral campaign in the context of Twitter: 2011 Spanish Presidential
election as a case study. Chaos 22(2):023138. https://doi.org/10.1063/
1.4729139
Bovet A, Makse HA (2018) Influence of fake news in Twitter during the 2016 US
presidential election. Nat Commun 10(1):7. http://arxiv.org/abs/1803.08491
Bovet A, Morone F, Makse HA (2018) Validation of Twitter opinion trends with
national polling aggregates: Hillary Clinton vs Donald Trump. Sci Rep 8
(1):8673. https://doi.org/10.1038/s41598-018-26951-y
Boyd D, Crawford K (2012) Critical questions for big data: provocations for a cultural,
technological, and scholarly phenomenon. Inf, Commun Soc 15(5):662–679
Bruns A, Stieglitz S (2012) Quantitative approaches to comparing communication
patterns on twitter. J Technol Hum Serv 30(3–4):160–185. https://doi.org/
10.1080/15228835.2012.744249
Caldarelli G (2010) Scale-Free Networks: Complex Webs in Nature and Tech-
nology. Oxford University Press, Oxford
Caldarelli G, Chessa A, Pammolli F, Pompa G, Puliga M, Riccaboni M, Riotta G
(2014) A multi-level geographical study of italian political elections from twitter
data. PLoS ONE 9(5):e95809. https://doi.org/10.1371/journal.pone.0095809
CiminiG, Squartini T, Saracco F, Garlaschelli D, Gabrielli A, Caldarelli G(2018)
The Statistical Physics of Real-World Networks Nat Rev Phys 1(1):58. http://
arxiv.org/abs/1810.05095
Conover M, Ratkiewicz J, Francisco M (2011) Political polarization on twitter.
Icwsm. https://doi.org/10.1021/ja202932e
Cram L, Llewellyn C, Hill R, Magdy W (2017) UK General Election 2017: a Twitter
Analysis. Neuropolitics Res Lab 2017:1–11. http://arxiv.org/abs/1706.02271
Dandekar P, Goel A, Lee D (2012) Biased Assimilation, Homophily and the
Dynamics of Polarization. Proc Natl Acad Sci USA 110(15):5791–5796.
https://doi.org/10.1073/pnas.1217220110
de Jeude J. v. L., Di Clemente R, Caldarelli G, Saracco F, Squartini T (2018)
Reconstructing mesoscale network structures. https://doi.org/10.1242/jeb.01477
Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley HE, Quat-
trociocchi W (2016a) The spreading of misinformation online. Proc Natl Acad
Sci 113(3):554–559. https://doi.org/10.1073/pnas.1517441113. ISSN 0027-8424

Del Vicario M, Vivaldo G, Bessi A, Zollo F, Scala A, Caldarelli G, Quattrociocchi W (2016b) Echo Chambers: Emotional Contagion and Group Polarization on Facebook. Sci Rep 6:37825. https://doi.org/10.1038/srep37825. ISSN 20452322

Del Vicario M, Gaito S, Quattrociocchi W, Zignani M, Zollo F (2017a) Public discourse and news consumption on online social media: a quantitative, cross-platform analysis of the Italian Referendum 10.475/123. http://arxiv.org/abs/1702.06016

Del Vicario M, Zollo F, Caldarelli G, Scala A, Quattrociocchi W (2017b) Mapping social dynamics on Facebook: The Brexit debate. Soc Netw 5:6–16. https://doi.org/10.1016/j.socnet.2017.02.002

Di Fonzo N (2011) The echo-chamber effect. https://www.nytimes.com/roomfordebate/2011/04/21/barack-obama-and-the-psychology-of-the-birther-myth/the-echo-chamber-effect

Di Gangi D, Lillo F, Pirino D (2018) Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. J Econ Dyn Control 94:117–141. https://doi.org/10.1016/j.jedc.2018.07.001

Diakopoulos NA, Shamma DA (2010) Characterizing debate performance via aggregated twitter sentiment. In: Proceedings of 28th 28th International Conference on Human Factors in Computing Systems. CHI '10. ACM Press, New York

Diestel R (2012) Graph theory, 4th edn. https://doi.org/10.1109/IEMBS.2010.5626521

DiGrazia J, McKelvey K, Bollen J, Rojas F (2013) More tweets, more votes: social media as a quantitative indicator of political behavior. PLoS ONE 8(11): e79449. https://doi.org/10.1371/journal.pone.0079449

Driscoll K, Walker S (2014) Big data, big questions| working within a black box: transparency in the collection and production of big twitter data. Int J Commun 8:20

Enli GS, Skogerbø E (2013) Personalized campaigns in party-centered politics: Twitter and Facebook as arenas for political communication. Inf Commun Soc 16(5):757–774. https://doi.org/10.1080/1369118X.2013.782330

Ferrara E (2017) Disinformation and social bot operations in the run up to the 2017 French presidential election. First Monday. https://doi.org/10.2139/ssrn.2995809

Ferrara E, Yang Z (2015) Quantifying the effect of sentiment on information diffusion in social media. PeerJ Comput Sci 1:e26. https://doi.org/10.7717/peerj-cs.26

Flaxman S, Goel S, Rao JM (2013) Ideological segregation and the effects of social media on news consumption. SSRN Electron J 1–39. https://doi.org/10.2139/ssrn.2363701

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Fronczak P, Fronczak A, Bujok M (2013) Exponential Random Graph Models for Networks with Community Structure. Phys Rev E 88(May):032810. https://doi.org/10.1103/PhysRevE.88.032810

Garlaschelli D, Loffredo MI (2008) Maximum likelihood: Extracting unbiased information from complex networks. Phys Rev E - Stat Nonlinear, Soft Matter Phys 78(1):1–5. https://doi.org/10.1103/PhysRevE.78.015101

Gibson RK, McAllister I (2006) Does Cyber-Campaigning Win Votes? Online Communication in the 2004 Australian Election. J Élect, Public Opin Parties 16(3):243–263. https://doi.org/10.1080/13689880600950527

González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. Sci Rep 1:197. https://doi.org/10.1038/srep00197

González-Bailón S, Borge-Holthoefer J, Moreno Y (2013) Broadcasters and Hidden Influentials in Online Protest Diffusion. Am Behav Sci 57(7):943–965. https://doi.org/10.1177/0002764213479371

González-Bailón S, Wang N, Rivero A, Borge-Holthoefer J, Moreno Y (2014) Assessing the bias in samples of large online networks. Soc Netw 38:16–27

Gualdi S, Cimini G, Primicerio K, Di Clemente R, Challet D (2016) Statistically validated network of portfolio overlaps and systemic risk. Sci Rep 6:39467. https://doi.org/10.1038/srep39467

Hampton KN, Shin I, Lu W (2017) Social media and political discussion: when online presence silences offline conversation. Inf, Commun Soc 20(7):1090–1107

Jaynes ET (1957) Information Theory and Statistical Mechanics. Phys Rev 106:620–630. https://doi.org/10.1103/PhysRev.106.620

Katz E, Lazarsfeld PF, Columbia University (1955) Personal influence: The part played by people in the flow of mass communications. Free Press, Glencoe, Ill

Kobayashi R, Lambiotte R (2016) Tideh: time-dependent hawkes process for predicting retweet dynamics. In: Proc. 10th International Conference on Web and Social Media, ICWSM 2016. AAAI press, Cambridge, MA, US

Lazarsfeld PF, Berelson B, Gaudet H (1944) The People's Choice. How the Voter Makes Up His Mind in a Presidential Campaign. Duell, Sloan and Pearce

Mastrandrea R, Squartini T, Fagiolo G, Garlaschelli D (2014) Enhanced reconstruction of weighted networks from strengths and degrees. New J Phys 16 (4):043022. https://doi.org/10.1088/1367-2630/16/4/043022

Morstatter F, Pfeffer J, Liu H, Carley KM (2013) Is the sample good enough? Comparing data from twitter's streaming API with Twitter's firehose. In: Proc. 7th International Conference on Weblogs and Social Media, ICWSM 2013. AAAI press. Cambridge, MA, US, p 400–408

Moy P, Domke D, Stamm K (2001) The spiral of silence and public opinion on affirmative action. J Mass Commun Quaterly 78(1):7–25. https://doi.org/10.1177/107769900107800102

Newman M (2010) Networks: an introduction. Oxford University Press, Inc. New York, NY, USA

Nikolov D, Oliveira DFM, Flammini A, Menczer F (2015) Measuring Online Social Bubbles. PeerJ Comput Sci 1:e38. https://doi.org/10.7717/peerj-cs.38

Oshagan H (1996) Reference group influence on opinion expression. Int J Public Opin Res 8(4):335–354. https://doi.org/10.1093/ijpor/8.4.335

Park J, Newman MEJ (2004) Statistical mechanics of networks. Phys Rev E 70 (6):66117. https://doi.org/10.1103/PhysRevE.70.066117

Pfeffer J, Mayer K, Morstatter F (2018) Tampering with twitter's sample api. EPJ Data Sci 7(1):50

Quattrociocchi W, Caldarelli G, Scala A (2014) Opinion dynamics on interacting networks: media competition and social influence. Sci Rep 4:4938. https://doi.org/10.1038/srep04938

Saracco F, Di Clemente R, Gabrielli A, Squartini T (2015) Randomizing bipartite networks: the case of the World Trade Web. Sci Rep 5:10595

Saracco F, Straka MJ, Di Clemente R, Gabrielli A, Caldarelli G, Squartini T (2017) Inferring monopartite projections of bipartite networks: An entropy-based approach. New J Phys 19(5):16. https://doi.org/10.1088/1367-2630/aa6b38

Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W (2018a) Polarization of the vaccination debate on Facebook. Vaccine 36(25):3606–3612. https://doi.org/10.1016/j.vaccine.2018.05.040

Schmidt AL, Zollo F, Scala A, Quattrociocchi W (2018b) Polarization rank: a study on European News Consumption on Facebook. http://arxiv.org/abs/1805.08030

Squartini T, Garlaschelli D (2011) Analytical maximum-likelihood method to detect patterns in real networks. New J Phys 13 https://doi.org/10.1088/1367-2630/13/8/083001

Squartini T, van Lelyveld I, Garlaschelli D (2013) Early-warning signals of topological collapse in interbank networks. Sci Rep 3:3357. https://doi.org/10.1038/srep03357

Squartini T, Garlaschelli D (2017) Maximum-entropy networks. Pattern detection, network reconstruction and graph combinatorics. Springer International Publishing, Switzerland, p 116

Stella M, Cristoforetti M, De Domenico M (2018a) Influence of augmented humans in online interactions during voting events. http://arxiv.org/abs/1803.08086

Stella M, Ferrara E, De Domenico M (2018b) Bots sustain and inflate striking opposition in online social systems. http://arxiv.org/abs/1802.07292

Tabourier L, Libert AS, Lambiotte R (2016) Predicting links in ego-networks using temporal information. EPJ Data Sci 5(1):1. https://doi.org/10.1140/epjds/s13688-015-0062-0

TNS Opinion and Social and Directorate-General Communications (2018) Media use in the European Union. https://publications.europa.eu/en/publication-detail/-/publication/a575c1c9-58b6-11e8-ab41-01aa75ed71a1/language-en/format-PDF/source-77908075

Varol O, Ferrara E, Menczer F, Flammini A (2017) Early detection of promoted campaigns on social media. EPJ Data Sci 6(1):13. https://doi.org/10.1140/epjds/s13688-017-0111-y

Acknowledgements

Author contributions
FS collected the data. All authors conceived the experiments, while CB and FS designed and performed the analyses and prepared tables and figures. All authors wrote, reviewed and approved the paper.

Additional information