# scientific reports

OPEN

# Analysing Twitter semantic networks: the case of 2018 Italian elections

Tommaso Radicioni[1,2✉], Fabio Saracco[2], Elena Pavan[3] & Tiziano Squartini[2]

Social media play a key role in shaping citizens' political opinion. According to the Eurobarometer, the percentage of EU citizens employing online social networks on a daily basis has increased from 18% in 2010 to 48% in 2019. The entwinement between social media and the unfolding of political dynamics has motivated the interest of researchers for the analysis of *users online behavior*—with particular emphasis on *group polarization* during debates and *echo-chambers formation*. In this context, *semantic aspects* have remained largely under-explored. In this paper, we aim at filling this gap by adopting a two-steps approach. First, we identify the *discursive communities* animating the political debate in the run up of the 2018 Italian Elections as groups of users with a significantly-similar retweeting behavior. Second, we study the mechanisms that shape their internal discussions by monitoring, on a daily basis, the structural evolution of the semantic networks they induce. Above and beyond specifying the semantic peculiarities of the Italian electoral competition, our approach innovates studies of online political discussions in two main ways. On the one hand, it grounds semantic analysis within users' behaviors by implementing a method, rooted in statistical theory, that guarantees that our inference of socio-semantic structures is not biased by any unsupported assumption about missing information; on the other, it is completely automated as it does not rest upon any manual labelling (either based on the users' features or on their sharing patterns). These elements make our method applicable to any Twitter discussion regardless of the language or the topic addressed.

In the last decade, social media platforms has brought fundamental changes to the way information is produced, communicated, distributed and consumed. According to Eurobarometer, the percentage of Europeans employing online social networks on a daily basis has increased from 18% in 2010 to 48% in 2019[1]. A similar report concerning the US showed that, as of August 2018, 68% of American adults retrieve at least some of their news on social media[2]. As social media facilitate rapid information sharing and large-scale information cascades, what emerges is a shift from a *mediated*, *top-down* communication model heavily ruled by legacy media to a *disintermediated*, *horizontal* one in which citizens actively select, share and contribute to the production of politically relevant news and information, in turn affecting the political life of their countries[3]. In a context in which political dynamics unfold with no solution of continuity within a hybrid socio-political media space, a multiplicity of studies that cut across traditional disciplinary boundaries have multiplied that uncover the many implications of users online behavior for political participation and democratic processes.

The systematic investigation of online networks spurring from social media use during relevant political events has been particularly helpful in this respect. Endorsing a view of online political activism as complementary to—and not as a substitution for—traditionally studied political participation dynamics[4], detailed and data-intensive explorations of online systems of interactions contributed to a more genuine and multilevel understanding of how social media relate to political participation processes.

At the macro level, research has focused on mapping the structural and processual features of online interaction systems to elaborate on the social media potential for fostering democratic and inclusive political debates. In this respect, specific attention has been paid to assessing grades of polarization and closure[5,6] of online discussions within echo-chambers[7] with a view of connecting such features with the progressive polarization of political dynamics[8,9].

At the micro level, attention has gone towards disambiguating the different roles that social media users may play—particularly, to identify influential spreaders[10–12] responsible for triggering the pervasive diffusion of certain types of information, but also to elaborate on the redefinition of political leadership in comparison to

[1]Scuola Normale Superiore, P.zza dei Cavalieri 7, 56126 Pisa, Italy. [2]IMT School for Advanced Studies, P.zza S. Ponziano 6, 55100 Lucca, Italy. [3]University of Trento, Via Verdi 26, 38122 Trento, Italy. ✉email: tommaso.radicioni@sns.it

more traditional offline dynamics[13]. More specifically, accounting for users behavior has helped characterizing the different contributions that are delivered by actors who exploit to different extents social media communication and networking potentials[14–16]. In this way, concepts like "political relevance" and "leadership" get redefined at the crossroads between actors' attributes and their actual engagement within online political discussions.

Additionally, increasing attention to online dynamics has entailed dealing with non-human actors, such as platform algorithms[17] and bots[18–21]. Consideration for non-human actors follows from extant social sciences approaches such as the actor-network theory that invites to disanchor agency from social actors, preferring a recognition for actants (i.e., any agent capable of intervening within social dynamics)[22]. Consistently, the pervasive diffusion of social media in every domain of human action revamps attention for both platform materiality (i.e. the modes in which specific technological artifacts are constructed and function) and for actants, starting from the premise that online dynamics are inherently sociotechnical and, thus, technology features stand in a mutual and co-creative relationship with their social understanding and uses[23]. Shrouded in invisibility, platform algorithms and social bots actively filter and/or push specific types of contents thus bending users behaviors and opinions—in some cases acting as true agents of misinformation[24,25].

In all its heterogeneity, this variety of studies shares a common feature, insofar as it mostly grounds in the study of *networks of users* and, thus, approaches the study of online political dynamics by privileging the investigation of direct relations amongst actors of different nature—individuals, organizations, institutions and even bots. Conversely, less attention has gone towards the *contents* that circulate during online political discussions and how these contribute to nurture collective political identities which, in turn, drive political action and participation.

To be sure,studies that focus on social media content do exist and embrace a multiplicity of political instances, from electoral campaigns to social movements and protests. For example, looking at Twitter, research has compared the content of tweets published by parties with the content of tweets sent by candidates[26], analyzed the contents of the 2017 French presidential electoral campaign[27], the online media coverage in the run up of the 2018 Italian Elections[28] and looked at the keywords and hashtags related to the *#MeToo* movement[29]. Nonetheless, when the focus has been set on social media contents, only rarely have these been investigated in connection with systems of social relations established amongst users on social media platforms [30,31]. Ultimately, the social and semantic aspects have hitherto been studied independently and we are still missing empirical pathways to explore the nexus between contents of online political conversations and the relational systems amongst users sustaining them.

This paper aims at filling in this gap by proposing an innovative approach that links the identification of communities of users that display a similar communication behavior with a thorough investigation of the most prominent contents they discuss. More specifically, by looking at the online conversation that unfolded on Twitter in the run up of the 2018 Italian Elections, our paper identifies *discursive communities* as *groups of users with a similar retweeting behavior* and analyses the contents circulating within these online political communities.

Several techniques of inferring the political alignment of Twitter users have been already implemented looking specifically at electoral campaigns, such as in the run-up to the 2010 U.S. midterm elections[32], the 2012 French presidential elections[33] and the 2016 U.S. elections[34]. Inference methods adopted in these works are based either on the construction of a retweet network based on the manual labelling of (at least) a fraction of Twitter users or on sharing patterns of web-links and/or hashtags. Moreover, the final number of communities is often defined a-priori, as every link or hashtag is assigned a specific political connotation linked to existing party systems or the left-right ideological spectrum.

Against this background, the approach we adopt here to infer discursive communities is innovative in two ways. First, it detects groups of Twitter users without any manual labelling (either based on users' features or on content sharing patterns). Second, the political valence of online actions such as retweeting is behavior-driven rather than ideologically determined as it is based on the direct action, performed by unverified users, of expanding the reach of verified users' messages via retweets.

Having identified communities in this way, we turn towards the study of contents they discuss by examining the semantic networks induced by the co-occurrences of hashtags contained in the tweets sent by their members. Through this procedure, it is possible to examine the semantic aspects of online political debates while grounding them in users' behaviors. Finally, we implement several filtering algorithms[35] to detect the non-trivial content of our semantic networks, identifying the most debated subjects. Filtering ultimately allows us to identify the communication strategies adopted by the different discursive communities and the backbone of the narratives developed by the different groups.

The paper is organized as follows. Section Case Study and Data describes data-acquisition and data-cleaning processes. In Section Methods, we discuss the methods we employ to project our bipartite user-hashtag networks on the hashtag layer and to derive our collection of semantic networks. We report and discuss results of our analysis of the Italian case in the Sections Results and Discussion. Finally, in Section Conclusions we further elaborate on the potentialities and limits of our proposed approach.

## Case study and data

**Case study.** The current study focuses on the Twitter discursive communities that emerged during the weeks of the electoral campaign preceding the Italian Elections on 4 March 2018. The 2018 Italian Elections represented a crucial political event that subverted the traditionally bipolar political competition characterizing the so-called Italian Second Republic. A radically novel scenario, with three poles of attraction did in fact emerge. The first pole was represented by the centre-right coalition which eventually won the elections with 37% of the vote share. Interestingly, the victory of the right-wing alliance was not led by Silvio Berlusconi's party, `Forza Italia', which obtained only a 14% of preferences and thus gave way to the nationalist `Lega' led by Matteo Salvini (17,4%). The second pole was represented by the center-left coalition led by `Partito Democratico' (PD) with

18.7% of the vote share—its worst result ever—under the leadership of the secretary and Prime Minister candidate Matteo Renzi. The third pole was represented by the populist party `Movimento 5 Stelle' (M5S) which unexpectedly obtained 32.7% of preferences.

Ultimately, the 2018 elections constituted a true electoral earthquake triggered by two elements. On the one hand, the extreme predominance of themes such as immigration and criminality which eventually favored populist and right-wing parties over more traditional actors such as `Forza Italia' and the `Partito Democratico'. On the other hand, a significant contribution to the shuffling of political balances was given also by the hybrid electoral campaign[36] put in place by all leaders and candidates who combined traditional and social media and managed to engage voters with pervasive and low-cost communication strategies.

**Social media platform and relations selection.** Twitter is hardly the only social media platform that hosted politically relevant discussions during the observation period, as all social media platforms play an increasingly relevant political role[37,38]. Nonetheless, extant studies show that Twitter is particularly prominent during electoral dynamics[39] as it is the platform used by the vast majority of public figures (e.g. political leaders, journalists, official media accounts, etc.) to provide visibility to their statements. More specifically, in the Italian context, Twitter is recognized to play an "agenda setting" effect on the country's mass media[40]. Hence, regardless of the fact that Twitter users are not representative of the Italian population, looking at the discursive communities that emerge on this platform allows us to look at a pivotal—albeit non representative—portion of political discussions that accompanied the electoral process.

Amongst all types of interaction modes featured by the platform, the current study grounds on retweets, which we understand as a baseline online relational mechanism that is particularly insightful when studying collective political identities. Indeed, as pointed out in[4], while mentions and replies in Twitter do sustain direct interaction and dialogue between users, retweets suggest a will to re-transmit contents produced by others. This, in turn, provides a more clear-cut indication of commonality and shared points of reference. Moreover, extant research suggests that retweets proxy the actual political alliances better than mentions and replies—as shown in[32], where authors conclude that the use of retweets was more relevant than that of mentions to grasp the bipartisan nature of online debates in the run up of the 2010 US midterm elections.

**Data collection.** The extraction of Twitter data has been performed by selecting a set of keywords linked to the Twitter discussion about 2018 Italian Elections. In particular, each collected tweet contains at least one of the following keywords: *elezioni*, *elezioni2018*, *4marzo*, *4marzo2018* (literally, *elections*, *elections2018*, *4march*, *4march2018*). Data collection has been realized by using the Twitter Search API across a period of 51 days—from 28 January 2018 to 19 March 2018—a time interval covering the entire period of the electoral campaign and the two weeks after the Election Day.

**Data cleaning.** The procedure described above led to a data set containing 1.2 millions of tweets, posted by 123.210 users (uniquely identified via their user ID). As in the Twitter environment hashtags play a central role, acting as *thematic tags* designated by the hash symbol #[41], we defined the nodes of our semantic networks as the hashtags extracted from the text of tweets. Thus, as a consequence, only tweets containing at least one hashtag have been retained. This procedure left us with $\simeq$ 38% of the original dataset. Notably, this result indicates that only a small percentage of users employs at least one hashtag while tweeting, as already reported elsewhere[42]. Hashtags were then subjected to a merging procedure: any two hashtags have been considered as the same if found "similar enough" and only the most frequent hashtag has been retained. The similarity between hashtags has been quantified through the *Levenshtein* or *edit distance* (see Supplementary Note 1 for more details), i.e. one of the most common sequence-based similarity measures[43]. This procedure allows us to avoid duplicated hashtags resulting from typing errors or linguistic variability. As shown by a check *a posteriori*, our cleaning procedure misidentifies less than 1% of the final list of hashtags.

**Data representation.** Then, we used the lists of user IDs and merged hashtags to define a bipartite network for each day of our observation period (51 bipartite networks in total). A bipartite network is defined by two distinct groups, or layers, of nodes, $\top$ and $\bot$, and only nodes belonging to different layers are allowed to be connected. The bipartite network corresponding to day $t$ can be, thus, represented as a matrix $\mathbf{M}^{(t)}$ whose dimensions are $N_\top \times N_\bot$, with $N_\top$ being the total number of users on day $t$ and $N_\bot$ being the total number of hashtags (tweeted) during that specific day: $m_{i\alpha}^{(t)} = 1$ if the user $i$ has tweeted (at least once) the hashtag $\alpha$ on day $t$ and 0 otherwise.

## Methods

The simplest way to obtain a monopartite projection out of a bipartite network is that of linking any two nodes belonging to the layer of interest (for example, $\alpha$ and $\beta$) if their number of common neighbors is *positive*. Such a procedure yields an $N_\bot \times N_\bot$ adjacency matrix $\mathbf{A}$ whose generic entry reads

$$a_{\alpha\beta} = \Theta[V_{\alpha\beta}^*] \tag{1}$$

where

$$V_{\alpha\beta}^* = \sum_{j=1}^{N_\top} m_{\alpha j} m_{\beta j} \tag{2}$$

counts the number of nodes both $\alpha$ and $\beta$ are linked to and $\Theta$ represents the Heaviside step function. The condition $a_{\alpha\beta} = \Theta[V^*_{\alpha\beta}] = 1$ can be also rephrased by saying that $\alpha$ and $\beta$ share *at least* one common neighbor.

A more refined method to obtain a monopartite projection is that of linking any two nodes if the number of common neighbors is found to be *statistically significant* according to an entropy-based framework[35]. More quantitatively, this second algorithm prescribes to compare the empirical value $V^*_{\alpha\beta} = \sum_{j=1}^{N_\top} m_{\alpha j} m_{\beta j}$ with the outcome of a properly-defined benchmark model—here, generically indicated with $f$—via the calculation of the p-value

$$\text{p-value}(V^*_{\alpha\beta}) = \sum_{V_{\alpha\beta} \geq V^*_{\alpha\beta}} f(V_{\alpha\beta}) \tag{3}$$

and link $\alpha$ and $\beta$ only in case this value "survives" a multiple hypotheses test (see Supplementary Note 2 for more details). Such a procedure outputs an $N_\perp \times N_\perp$ adjacency matrix **A** whose generic entry reads $a_{\alpha\beta} = 1$ if nodes $\alpha$ and $\beta$ are found to be linked to the same neighbors a statistically significant number of times and $a_{\alpha\beta} = 0$ otherwise.

The null models used as filters for our analysis are the *Bipartite Random Graph Model* (BiRGM), the *Bipartite Partial Configuration Model* (BiPCM) and the *Bipartite Configuration Model* (BiCM)[35] (see Supplementary Note 2 for more details). In a nutshell, the BiRGM discounts the information provided by the total number of (re) tweets, the BiPCM discounts the information provided by the total number of (re)tweeted hashtags per user, and the BiCM discounts the information provided by both the total number of (re)tweeted hashtags per user and the total number of (re)tweeting users per hashtag.

For every bipartite network in our data set, we created two matrices following the outlined procedure: first, a monopartite user by user network that we employed to identify discursive communities; second, a monopartite hashtag by hashtag network that we employed to study the contents discussed within the identified discursive communities.

## Results

### User by user networks and discursive communities.

Our first step to analyze the Twitter public discourse of the 2018 Italian electoral campaign consists of identifying communities of online users with a similar behavior. To this aim, we have divided users into two groups, by distinguishing the accounts *verified* by the platform from the *non-verified* ones. It is worth noting that the account verification procedure can be requested by any user to guarantee to other Twitter users that the account is authentic: for this reason, the verified accounts are usually composed by "entities" such as politicians, journalists, political parties or media. This information can be easily retrieved by employing the Twitter APIs. A bipartite network is, then, built as follows: a verified and a non-verified user are linked if one of the two retweets the other at least once during the observation period—notably, the retweeting action is mainly performed by non-verified users who share contents published by the verified ones. Then, the procedure described in Section Methods has been employed to project the bipartite network of retweets on the layer of verified users employing the BiCM filter. Lastly, a traditional community detection algorithm has been run to identify communities of verified users (see Supplementary Note 3 for more details). These groups constitute *discursive communities* wherein the tweeting activity of the verified users triggers a discussion between the non-verified users sharing similar contents. Moreover, a handful of non-verified users are assigned to the communities of the verified ones, via the computation of the so-called *polarization* (see Supplementary Note 4 and also[11,44]).

Interestingly, identified discursive communities provide a faithful representation of the alliances running at the 2018 Italian Elections and of their supporters:

- 'Movimento 5 Stelle' (M5S): a community composed by accounts of politicians belonging to 'Movimento 5 Stelle' (e.g. *DaniloToninelli*, *luigidimaio*), institutional accounts of the party (e.g. *M5S_Camera*, *M5S_Senato*) and users engaging with all of them. The number of users belonging to this community is 11.151;
- Center-right (CDX): a community of users composed by accounts of allied right-wing political parties (e.g. *forza_italia*, *LegaSalvini*), center and right-wing wing politicians (e.g. *renatobrunetta*, *matteosalvinimi*), their institutional representative groups (e.g. *GruppoFICamera*), and users interacting with all of these. The number of users belonging to this community is 5.842;
- Center-left (CSX): a rather heterogeneous community of users composed by accounts of political parties composing the center-left alliance (e.g. *pdnetwork*, *PD_ROMA*), their politicians (e.g. *giorgio_gori*, *matteorenzi*), some journalists (e.g. *vittoriozucconi*, *jacopo_iaboni*), and users engaging with them. The number of users belonging to this community is 12.065.

### Activity level of discursive communities.

A first step in the analysis of these discursive communities consists of examining their volume of activity. As Fig. 1 shows, the evolution of the Twitter activity of the three discursive communities is similar. Generally speaking, a flat trend is followed by a steep rise few days before the Election Day; then, a peak in the tweeting activity is registered the day *after* elections (5 March 2018). Afterwards, a rapid decrease of the number of tweets can be observed: in comparison with values preceding the Election Day, the volume of CDX tweets decreases by $\simeq 60\%$, the volume of M5S tweets decreases by $\simeq 50\%$ and the volume of CSX tweets decreases by $\simeq 20\%$. It is worth noting that the volume of tweets characterizing the M5S community is systematically larger than the volume of tweets characterizing both the CDX and the CSX community across the entire period considered here, an element that confirms the renowned attitude of M5S supporters towards the use of digital media for their organizational and public communication activities.
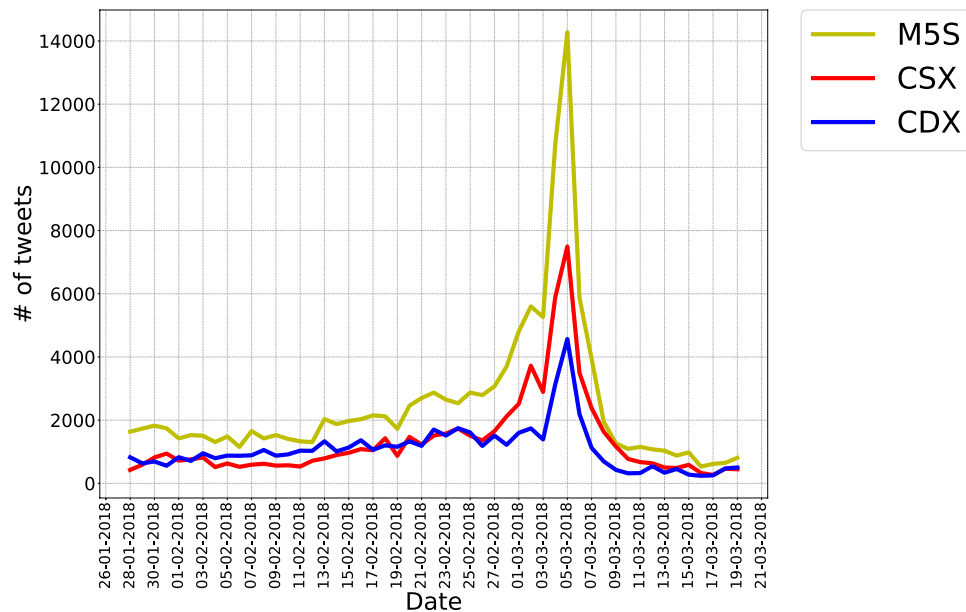
**Figure 1.** Volume of tweets characterizing the M5S, CDX and CSX communities across the observation period: notice the peak of activity, evident for all communities, registered in correspondence of the day *after* elections (5 March 2018). The volume of tweets characterizing the M5S community is systematically larger than the volume of tweets characterizing the other two communities, across the entire period considered here—an element that confirms the attitude of M5S supporters to use social media in political ways.

**Hashtag by hashtag networks.** Let us now move to the analysis of the monopartite projections on the layer of hashtags which we label as *hashtag by hashtag* or *semantic networks*. Below, we discuss the results generated by non-filtered projections. In the next section, we compare these results with those generated via filtered projections.

*Analyzing the topics prominence.* A closer inspection of semantic networks allows us to engage more systematically with the contents discussed within discursive communities. A first step in this direction can be made by exploring the number of *nodes*, which proxies the number of topics discussed by users, and their *mean degree* (i.e. the mean number of neighbors per node), which proxies the (average) prominence of the topics that characterize the discussion. Results obtained in this step are shown in Fig. 2.

The number of nodes shows a rising trend up to the day *after* elections, followed by a decreasing one. This indicates that the number of topics debated by users increases as 4 March 2018 approaches. Again, the M5S seems to be the most active community with the largest number of debated topics throughout our observation period. The trend characterizing the M5S community is closely followed by that of the right-wing alliance up to the end of February, when an inversion takes place and a rise in the number of topics debated by the supporters of the center-left alliance becomes clearly visible.

The trend of the mean degree is, overall, much less regular: it is, in fact, characterized by several 'bumps of activity' throughout the entire period. Peaks in the daily use of hashtags correspond to so-called *mediated events*, i.e., events of social relevance broadcast by mass media and, in particular, by national television channels. The importance of media events is suggested by the prominence of hashtags referring to Italian political talk shows, such as *#dallavostraparte* and *#tagadala7*. To confirm this intuition, tweets contributing to activity bumps have been qualitatively explored and, indeed, contents tweeted by users and mainstraem media have been to systematically match, particularly, in the case of TV shows featuring prominent politicians. We further cross-checked the contents of these tweets particularly with printed and online news to verify the actual presence of political leaders and figures during talk shows mentioned by users in their tweets. After this qualitative validation, we concluded that users tend to become particularly active during political debates signalling and/or commenting on the presence of certain candidates in television. Such a behavior is particularly evident for the CDX community, whose mean degree is characterized by a larger number of peaks. More specifically, the peaks are observed in correspondence of the following TV shows:

- 09 February: interview of Silvio Berlusconi at *TG La7* (hashtags: *#silvioberlusconi*, *#tgla7*);
- 11 February: Silvio Berlusconi and Matteo Salvini are interviewed at *Mezz'Ora In Più* (hashtags: *#il4marzo-votaefaivotareforzaitalia*, *#mezzorainpiù*);
- 13 February: Nicola Porro, an italian journalist, announces via a Facebook video, the topics that will be discussed on his TV show *Matrix*, broadcast by *Canale 5*, a TV channel owned by the Berlusconi family (hashtags: *#nicolaporro*, *#matrix*);
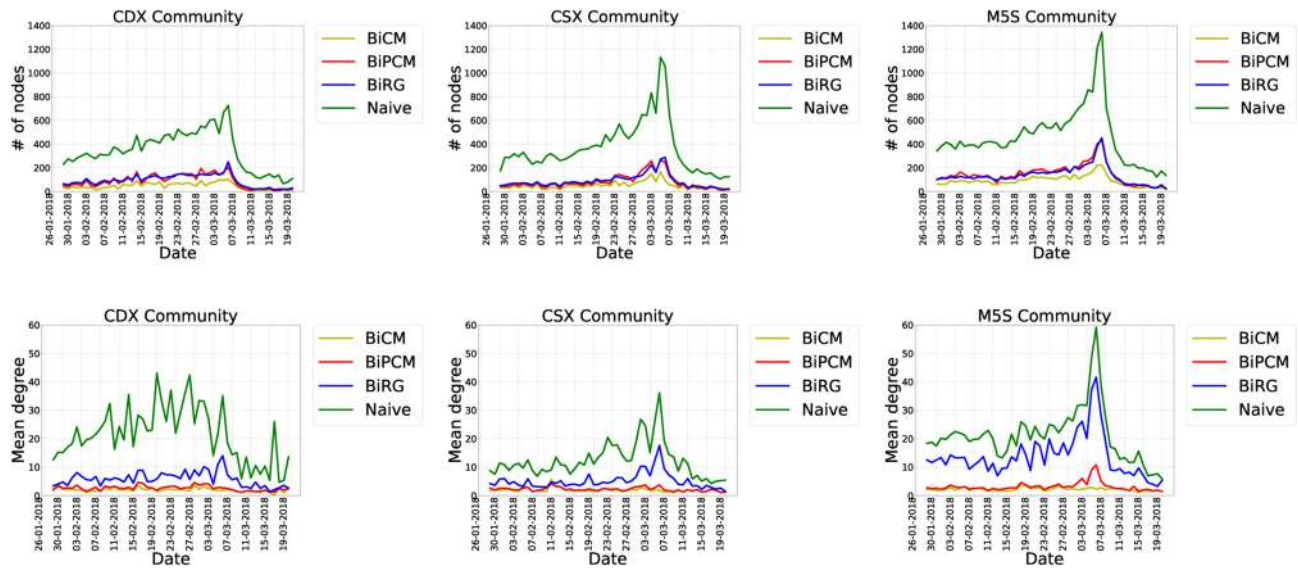
**Figure 2.** Temporal evolution of the number of nodes (top panels) and of the mean degree (bottom panels) for each community-specific semantic network. Peaks in the daily use of hashtags correspond to *mediated events*, i.e. events of social relevance broadcast by media. This behavior is particularly evident for the CDX community whose activity increases in correspondence with TV shows where politicians from the right-wing alliance are hosted.

- 18 February: interview of Silvio Berlusconi in the TV show *Che Tempo Che Fa*(hashtags: *#chetempochefa*, *#silvioberlusconi*);
- 19 February: interview of Silvio Berlusconi in the TV show *Dalla Vostra Parte*(hashtags: *#dallavostraparte*, *#silvioberlusconi*);
- 22 February: Matteo Salvini and Anna Maria Bernini (from the 'Forza Italia' party) are hosted in the TV show *Quinta Colonna* broadcast on *Rete 4*, another TV channel owned by the Berlusconi family (hashtags: *#forzaitaliaberlusconipresidente*, *#quintacolonna*);
- 26 February: Guido Crosetto and Maurizio Gasparri (both from the right-wing alliance) are hosted in the TV show *L'Aria Che Tira* (hashtag: *#lariachetirala7*);
- 16 March: interview of Michaela Biancofiore (from 'Forza Italia') in the TV show *Tagadà* (hashtags: *#tagada*, *#tagadala7*).

Beside confirming that Twitter discussions can be *influenced* by external events, our results point out that Twitter discussions can be also *triggered* by external events. This is especially true for the CDX community whose Twitter discussions do not emerge 'spontaneously' but are driven by the aforementioned mediated events[45], seemingly indicating that CDX users still conceive of television as the main information tool when it comes to political processes.

**Identifying persistent topics.** A second step towards a closer understanding of the contents discussed within discursive communities consists of quantifying the *interest towards a topic* throughout the entire period covered by our data set. To this aim, we analyzed *hashtag persistence*, $H_t$, i.e. the percentage of days an hashtag is present in our data set. Results are reported in Table 1. As it shows, the most persistent hashtags (in fact, the ones that are always present) are those concerning the name of political parties (i.e. *#lega*, *#m5s*, *#pd*) and political leaders (i.e. *#berlusconi*, *#dimaio*, *#renzi*, *#salvini*). Moreover, more persistent hashtags in all discursive communities refer almost in all cases to political actors and figures, more often than not of an opposing alliance. When it comes to substantive electoral themes, instead, the three communities seem to hold a common interest for work-related matters but also concentrate on peculiar interests: migration flows for the M5S, taxation for the CDX and the role of Europe for the CSX. This finding has been observed for all discursive communities and it highlights the fact that the online political debate largely focuses on single personalities/political entities and, albeit to a lesser extent, on themes of public interest.

**Identifying central topics.** In order to identify topics that, regardless of their prominence and persistence, are more pivotal to the unfolding of the discussion, we computed *hashtag betweenness centrality*, a measure quantifying the percentage of shortest paths passing through each hashtag, i.e.

$$b_\gamma = \sum_{\beta(\neq\alpha)} \sum_\alpha \frac{\sigma_\gamma^{\alpha\beta}}{\sigma^{\alpha\beta}} \tag{4}$$

| $H_t$ | M5S | CDX | CSX |
|---|---|---|---|
| 100% | dimaio, lega, renzi, berlusconi, m5s, pd, italia | salvini, m5s, centrodestra, pd, lega | renzi, salvini, dimaio, m5s, pd |
| 98% | forzaitalia, salvini | berlusconi, italia, renzi | |
| 96% | roma, ottoemezzo | forzaitalia | berlusconi, italia, lega |
| 94% | centrodestra, ricercapubblica | | russia |
| 92% | boschi, politica | dimaio | europa, politica, roma |
| 90% | fi, governo | fi, governo | |
| 88% | casapound | roma | |
| 86% | meloni | | |
| 84% | fakenews, lavoro, liberieuguali | casapound, politica | forzaitalia, lavoro, usa |
| 82% | 8800precari, gentiloni, migranti, senato, voto | governo, lombardia | centrodestra, leu, liberieuguali |
| 80% | bonino, campagnaelettorale, casini, leu, rosatellum | cdx, flattax, sinistra | milano, partitodemocratico, ue |
| 78% | avanti, iovotom5s, movimento5stelle, precari, sinistra | lavoro, ue | campagnaelettorale, fakenews, governo |

**Table 1.** Hashtag persistence for each discursive community across the entire temporal period covered by our data set (51 days in total). The first column shows the percentage of days each hashtag is present in the set of tweets of each community. Notice that the hashtags that are always present are those carrying the name of political parties and political leaders, while other relevant themes for the political debate are absent from (at least) some of the discursive communities. These findings suggest that the online political debate is largely focused on single personalities/political entities (as particularly evident upon inspecting the CSX hashtags) and only to a much smaller extent on themes of public interest.

(where $\sigma_\gamma^{\alpha\beta}$ is the number of shortest paths between hashtags $\alpha$ and $\beta$ passing through hashtag $\gamma$ and $\sigma^{\alpha\beta}$ is the total number of shortest paths between hashtags $\alpha$ and $\beta$). In this sense, hashtag betweenness centrality provides an entry point to identify strategic topics that "coordinate" the discussion, as they bridge other topics that users do not directly connect within their tweets. Interestingly, the basket of the most strategic hashtags (i.e. *#pd*, *#m5s*, *#renzi*, *#salvini*, *#berlusconi*, *#italia*, *#dimaio*, *#lega*, *#centrodestra*) is basically the same for all communities. This result reveals how the overall discussion is highly personalized as the main players of the 2018 Italian Elections embody crucial concepts for the definition of the narratives shaping the political debates of *all* communities. Nonetheless, the specificities of each community are maintained when it comes to economic and societal issues.

**Analysis of triadic closures.** As discussions develop around groups of hashtags, increasingly complex semantic structures are to be considered. To this aim, we analyzed the presence and the persistence of *triadic closures*, i.e. *triangles of connected hashtags*, which in our approach represent the core of larger semantic networks. In other words, they are the "seeds" around which more complex discussions grow—exactly as triangular motifs represent the simplest (yet informative) example of communities[46].

As it has been noticed, this kind of structure provides deeper insights into users' tweeting behavior, by revealing which concepts appear *simultaneously* in a discussion and measuring how often they do so[47]. This analysis is particularly insightful to distinguish the behavior of the three communities: as shown in Table 2, while both the CDX and the CSX communities are characterized by triads of concepts exclusively about political leaders, parties and electoral slogans, the triads observed within the M5S community reveal a greater concern of their members for themes of public interest (e.g. the issues of precarious labour, migrants landing, public research).

Interestingly, we also notice that specific days exist in which an abundance of triadic closures is registered. For instance, on the first day of the electoral silence, i.e. 2 March 2018, users are particularly active in building narratives around electoral slogans, while themes of public interest constitute the topic of tweets at the end of the electoral campaign (i.e. the last days of February). Finally, we notice that the abundance of hashtag triads tends to rise in correspondence with mediated events, as observed for the mean degree: this is the case for the days 27 February 2018 for the M5S community (when Luigi Di Maio was interviewed at the political talk show DiMartedì), 20 February 2018 for the CDX community (when Silvio Berlusconi was interviewed in a talk show called *#Italia18* organized by the Italian newspaper *Corriere della Sera*) and 23 February 2018 for the CSX community (when Laura Boldrini was interviewed at the radio show Circo Massimo).

**Analysis of degree-degree correlations.** A closer inspection of correlations between hashtags degrees allows us to elaborate more in depth on the ways prominent topics are connected to others, shaping broader politically relevant narratives. To this aim, we consider the *average nearest-neighbors degree* (ANND), defined, for the generic hashtag $\alpha$, as the arithmetic mean of the degrees of the neighbors of a node, i.e.

$$\kappa_\alpha^{nn} = \frac{\sum_{\beta(\neq\alpha)} a_{\alpha\beta}\kappa_\beta}{\kappa_\alpha}, \quad \forall\,\alpha \tag{5}$$

with $\kappa_\alpha$ indicating the degree of the hashtag $\alpha$ in the considered monopartite projection. The degree-degree correlation structure of a network can be easily inspected by plotting the $\kappa_\alpha^{nn}$ values versus the $\kappa_\alpha$ values. A *decreasing trend* would lead one to conclude that correlations between degrees are *negative*—that is, nodes with a small

| $T_t$ | M5S | CDX | CSX |
|---|---|---|---|
| 31% | (ricercapubblica, 8800precari, campagnaelettorale) | | |
| 27% | | | (salvini, pd, m5s) |
| 24% | (pd, italia, m5s) | | (pd, lega, m5s); (pd, dimaio, m5s) |
| 21% | (cnr, campagnaelettorale, ricercapubblica); (precari, campagnaelettorale, ricercapubblica); (politica, pd, m5s); (dimaio, pd, m5s); (lega, pd, m5s) | | (m5s, dimaio, salvini); (liberieuguali, pd,m5s); (m5s, berlusconi, pd); (usa, europa, russia); (savona, accettolasfida, poterealpopolo) |
| 20% | (berlusconi, pd, m5s); (ottoemezzo, pd, m5s); (salvini, pd, m5s); (berlusconi, politica, m5s); (centrodestra, pd, m5s); (italia, stopinvasione, italiani); (italia, stopislam, italiani); (campagnaelettorale, piemonte, forza-italia); (m5s, pd, m5salgoverno) | (salvini, pd, m5s) | (pd, m5s, renzi); (pd, italia, m5s); (salvini, lega, m5s); (forzaitalia, pd, m5s); (fattinonparole, partitodemocratico, avanti); (berlusconi, salvini, pd); (salvini, m5s, berlusconi) |
| **Dates** | | | |
| | 02 Mar 2018 | 20 Feb 2018 | 02 Mar 2018 |
| | 20 Feb 2018 | 27 Feb 2018 | 06 Mar 2018 |
| | 21 Feb 2018 | 02 Mar 2018 | 23 Feb 2018 |
| | 07 Mar 2018 | 01 Mar 2018 | 04 Mar 2018 |
| | 16 Feb 2018 | 22 Feb 2018 | 21 Feb 2018 |

**Table 2.** Persistence of triadic closures for each discursive community across the entire temporal period covered by our data set (51 days in total), on non-filtered projections. This analysis is particularly insightful to distinguish the behavior of the three communities: while the CSX and CDX communities are characterized by triads exclusively about political leaders, parties and electoral slogans, the triads observed within the M5S community focus more on themes of public interest. Notice that the largest $T_t$ value, i.e. the largest percentage of days a specific triadic closure is present in our data set, is sensibly less than the number of days covered by our data set (i.e. 51). Dates refer to the days with the largest number of triadic closures.

degree would be "preferentially" connected to nodes with high degree and vice versa. Conversely, an *increasing trend* would signal that correlations between nodes are *positive*—that is, nodes with a small (large) degree would be "preferentially" connected to nodes with a small (large) degree. Thus, decreasing and increasing trends offer us an entry point to explore whether discussions in the three communities tend to anchor to some key themes that work as conversational drivers.

The decreasing behavior of the ANND throughout our data set confirms the presence of negative degree-degree correlations, i.e. the considered networks are *disassortative* (less prominent hashtags are connected with more prominent hashtags and vice versa). Examples of these trends are reported in Fig. 3. The days considered here, i.e. 19 February 2018 and 5 March 2018, have been chosen to highlight an interesting feature of our semantic networks: as it is clearly visible upon inspecting the behavior of the CDX and the CSX communities, groups of nodes with a much larger value of the ANND appear. As it will become evident in what follows, these hashtags constitute the *core* of the Twitter discussion in the corresponding community and are characterized by a dynamics on a daily time-scale, i.e. they appear in correspondence of a specific event (in the case of the CDX community, the interview of Silvio Berlusconi in a TV show; in the case of the CSX community, Laura Boldrini's Twitter campaign) and disappear the day after.

As an additional analysis, we also considered the *clustering coefficient*, defined as

$$c_\alpha = \frac{\sum_{\gamma(\neq\alpha,\beta)}\sum_{\beta(\neq\alpha)} a_{\alpha\beta}a_{\beta\gamma}a_{\gamma\alpha}}{\kappa_\alpha(\kappa_\alpha - 1)}, \quad \forall\,\alpha \tag{6}$$

and quantifying the percentage of neighbours of a given node $\alpha$ that are also neighbours of each other (i.e. the percentage of triangles, having $\alpha$ as a vertex, that are actually present). As shown in Fig. 4, decreasing trends are observed: poorly-connected hashtags are strongly inter-connected and vice versa, thus suggesting the presence of several inter-connected "small" discussions that are connected to a group of central topics. A network with these features is also said to be *hierarchical*. Moreover, the hashtags with a larger value of ANND are also those with a larger value of the clustering coefficient—confirming the "coreness" of this group of topics. Taken altogether, these results suggest that all discursive communities revolve around a handful of few conversational drivers: overshadowed by the predominance of these issues, a set of niche discussions tend nonetheless to emerge, pointing out a variety of interests even within a single discursive community.

### Semantic networks at the mesoscale: k-core decomposition.

Shifting perspective onto the mesoscale structure of semantic network helps us to better clarify the power of conversational drivers we just identified. If triadic closure allowed us to identify the seeds to thematic discussions within communities, broadening the scope of the semantic analysis to the level of clusters allows in fact to explore the main conversational lines within the different communities.

In what follows, we focus our attention on 19 February 2018, but similar considerations hold true for other daily semantic networks. We implement the so-called *k-core decomposition*, a technique which has been widely used to find the structural properties of networks across a broad range of disciplines including ecology, economics and social sciences[48]. The k-core decomposition can be described as a sort of pruning process, where the nodes that have degree less than $k$ are removed, in order to identify the largest subgraph of a network whose nodes
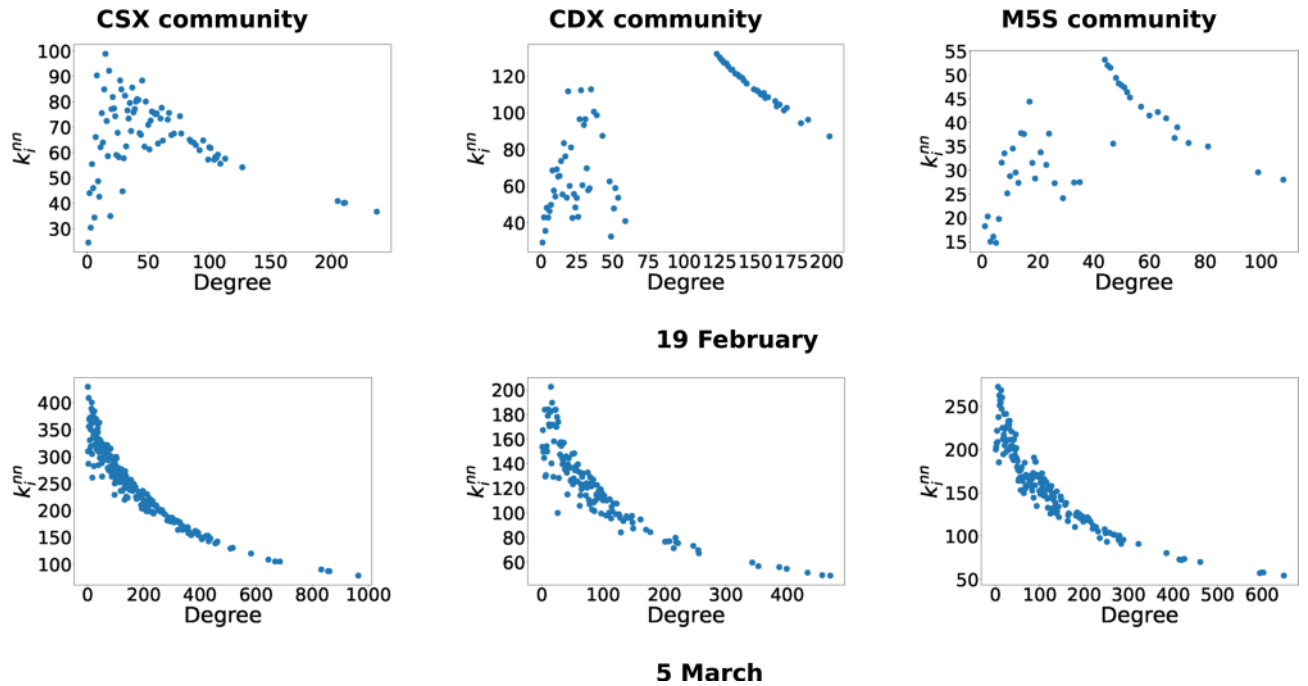
**Figure 3.** Analysis of the degree-degree correlations for two specific days, i.e. 19 February and 5 March 2018. As the trend of $\kappa_\alpha^{nn}$ reveals, the daily semantic networks are disassortative for all communities, i.e. nodes with small degree are (preferentially) connected to nodes with high degree and vice versa. A close inspection of behaviors of the CDX and the CSX communities shows groups of nodes with a larger value of the ANND: these clusters of hashtags constitute the core of the Twitter discussion in the corresponding community which appear in correspondence of specific events and disappear the day after.
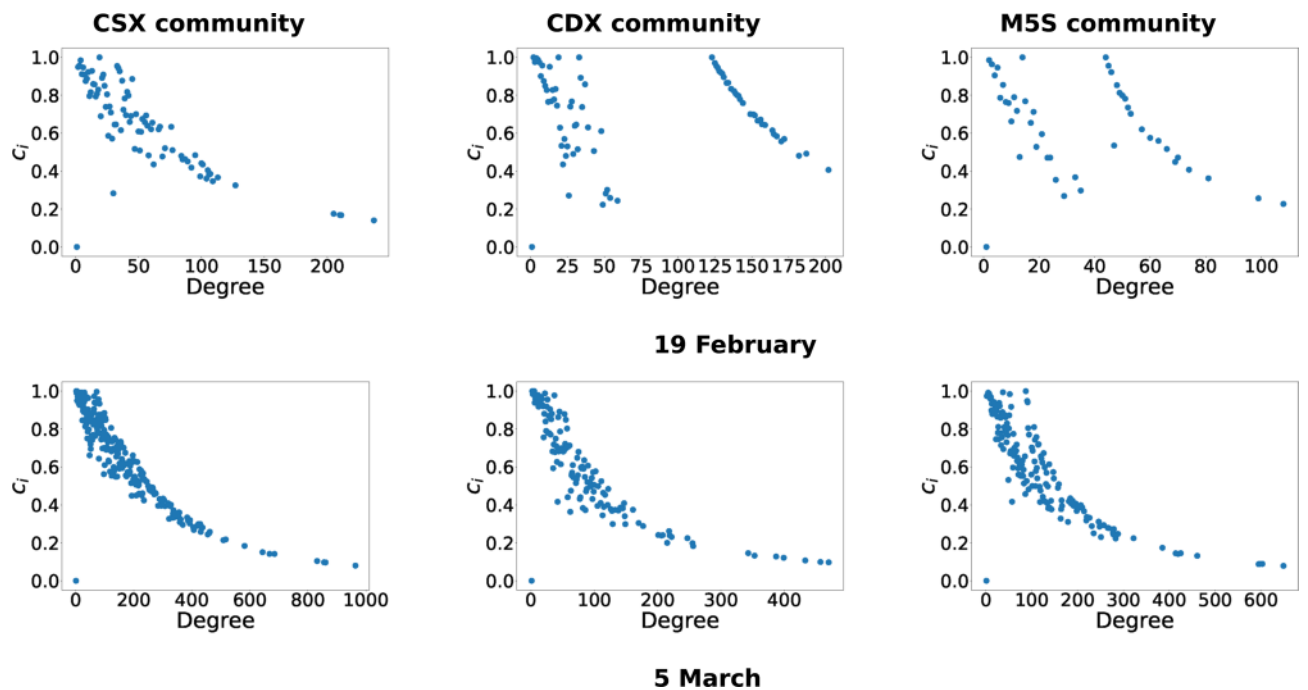


**Figure 4.** Analysis of network hierarchical structure for two specific days, i.e. 19 February 2018 and 5 March 2018. Plotting the clustering coefficient $c_\alpha$ values versus the degree $\kappa_\alpha$ values for the three communities reveals that our daily semantic networks are hierarchical, i.e. poorly-connected hashtags are strongly interconnected and vice versa. Besides, it also shows that the nodes with a larger value of the ANND are also the ones characterized by a larger value of the clustering coefficient.
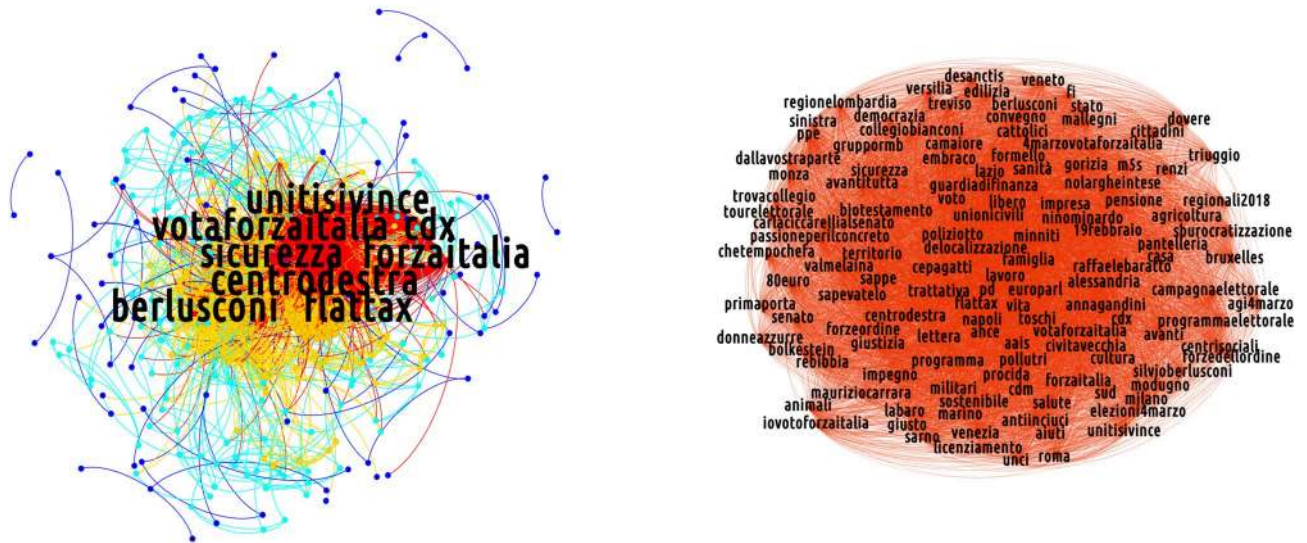
**Figure 5.** K-core decomposition of the semantic network for the non-filtered projection of the CDX discursive community on 19 February 2018. In the left plot, five k-shells are represented with different colors. In the right plot, an expanded view of the innermost k-shell—basically overlapping with the properly defined core identified by the bimodular surprise—is represented. The compact bulk is triggered by the interview of Silvio Berlusconi in the TV show *Dalla Vostra Parte*. [Picture realized with Gephi software version 0.9.2. For more information about Gephi[49]: https://gephi.org/].

have *at least* k neighbors. This method allows a "coreness" score to be assigned to each node of the network which remains naturally divided into shells. Node coreness is equal to k if the node is present in the k-core of the network but not in the $(k + 1)$-core.

Figures 5, 6 and 7 show the the k-shell decomposition for the semantic networks of our discursive communities, for the day 19 February 2018: five k-shells, corresponding to five quantiles of the degree distribution, have been colored, confirming the presence of a core of highly debated hashtags (the red one collecting the most prominent and intertwined ones). To inspect the presence of a substructure nested into the discussion core, we run the Louvain algorithm on the innermost k-shell of the semantic networks of our discursive communities. Their shell structure is indeed rich, as particularly evident when considering the CSX and the M5S ones: indeed, several communities appear, seemingly indicating that the discussions in which members of the two groups are (more) engaged self-organize around sub-topics.
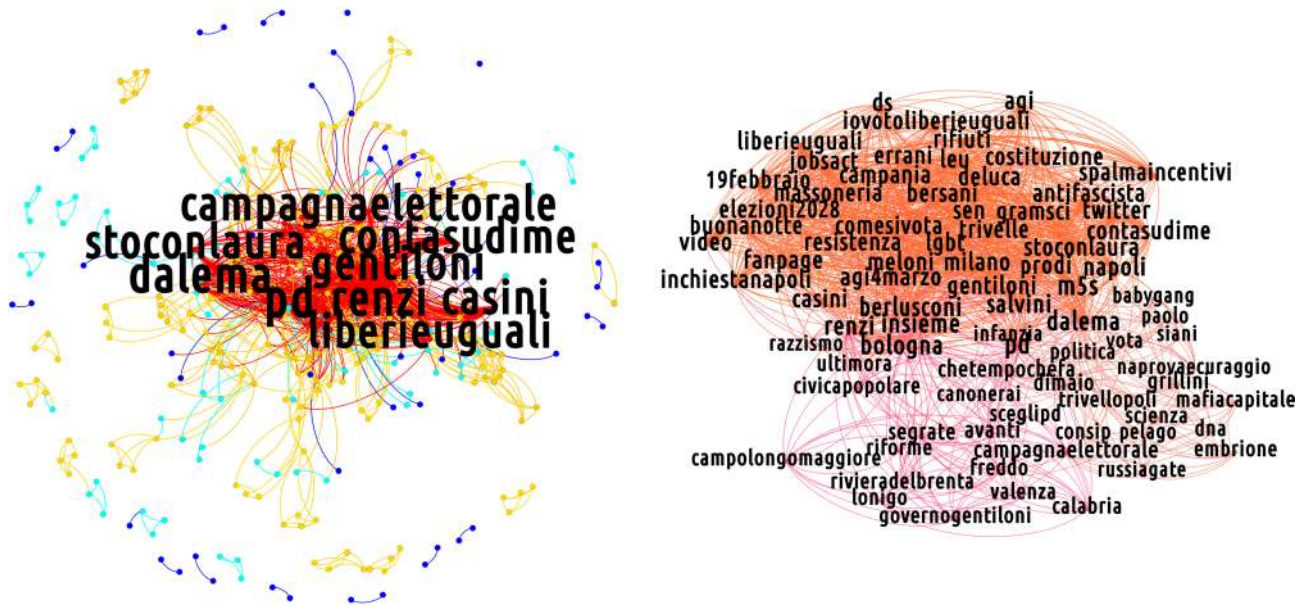
For what concerns the CSX community, the hashtags sub-communities emerge as a consequence of factors as the Twitter campaign born in support of the center-left candidate Laura Boldrini (revealed by the presence of hashtags such as *#stoconlaura* and *#contasudime*), the visit of Matteo Renzi in Bologna (revealed by the presence of hashtags such as *#bologna*, *#renzi*, *#errani*, *#casini*, hashtags that refer to Vasco Errani and Pier Ferdinando Casini, center-left wing candidates for the Senate in Emilia-Romagna) and the presence of Massimo D'Alema (another leader of the center-left alliance) in the radio show 'Circo Massimo' (revealed by the presence of hashtags such as *#dalema*).

On the other hand, the presence of multiple debates within the core of the M5S semantic network is related to events like the electoral tour of Alessandro Di Battista, a prominent figure of the party, who presented the M5S electoral program in the southern Italy region named Basilicata (hashtags: *#dibattista*, *#ilfuturoinprogramma*, *#programmaindiretta*, *#basilicata*), the presence of a journalist of *Il Fatto Quotidiano* (a newspaper supporting the M5S) invited in the TV show *Otto e Mezzo* (hashtags: *#ilfattoquotidiano*, *#ottoemezzo*) and the presence of politicians supporting other coalitions in several TV shows such as *Porta a Porta*, *Mezz'Ora In Più* and *Dalla Vostra Parte*.

However, these observations do not hold true when the CDX-induced semantic network is considered. Its innermost shell is, in fact, a compact group of topics that cannot be further partitioned.

As a second observation, we notice that—when present—the communities partitioning the core are held together by the nodes with largest betweenness centrality: as they coincide with the hashtags related to the name of political parties/leaders, the latter ones can be imagined to act as bridges connecting different discussions. Generally speaking, this indicates that the concept of "most influential nodes" can be applied also within the core of the networks of hashtags—a result that complements the one about the influential spreaders individuated within the networks of users[50].

**Semantic networks at the mesoscale: the core-periphery structure.** In order to complement the analysis above, we also implemented the method proposed in[51] which prescribe to search for the network core-periphery partition minimizing the quantity called *bimodular surprise*, i.e.

10

**Figure 6.** K-core decomposition of the semantic network for the non-filtered projection of the CSX discursive community on 19 February 2018. In the left plot, five k-shells are represented with different colors. In the right plot, an expanded view of the innermost k-shell—basically overlapping with the properly defined core identified by the bimodular surprise—is represented. Notice the presence of communities, found through the Louvain algorithm and emerging as a consequence of factors as diverse as the Twitter campaign born in support of the center-left candidate Laura Boldrini, the visit of Matteo Renzi in Bologna, the presence of Massimo D'Alema (another leader of the center-left alliance) in the radio show 'Circo Massimo'. (Picture realized with Gephi software version 0.9.2. For more information about Gephi[49]: https://gephi.org/).
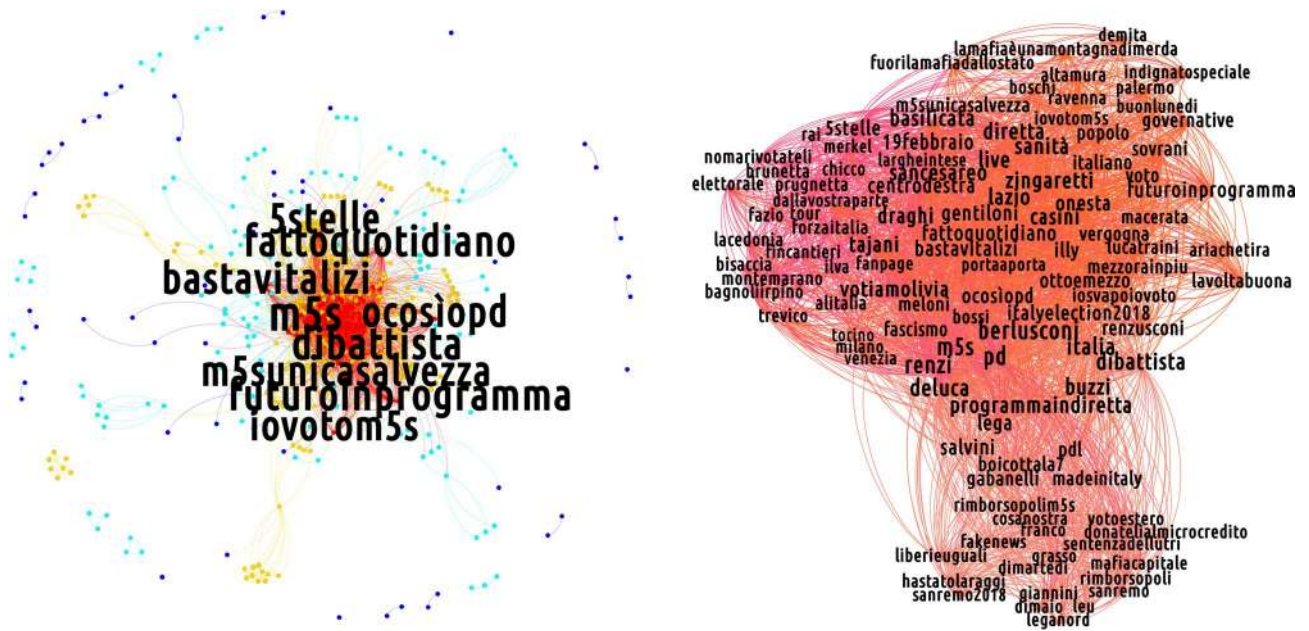


**Figure 7.** K-core decomposition of the semantic network for the non-filtered projection of the M5S discursive community on 19 February 2018. In the left plot, five k-shells are represented with different colors. In the right plot, an expanded view of the innermost k-shell—basically overlapping with the properly defined core identified by the bimodular surprise—is represented. Notice the presence of communities, found by running the Louvain algorithm. These emerge as a consequence of events as the electoral tour of Alessandro Di Battista (one of the M5S leaders), the presence of politicians in TV shows such as 'Porta a Porta', 'Mezz'ora in piú' and 'Dalla vostra parte'. [Picture realized with Gephi software version 0.9.2. For more information about Gephi[49]: https://gephi.org/].
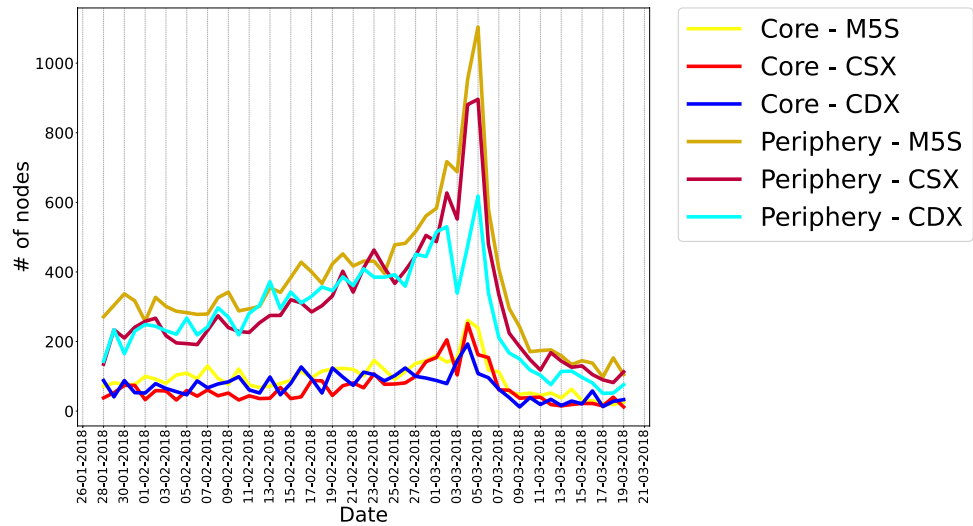
**Figure 8.** Evolution of the number of nodes belonging to the core and to the periphery of each discursive community. The core size is nearly constant throughout the observation period while the periphery size rises as the Election Day approaches (the peak appears in correspondence of the day after, i.e. the 5 March 2018). This behavior, common to all communities, reveals that, as the Election Day approaches, the number of topics animating the discussion increases.

$$\mathscr{S}_{\parallel} = \sum_{i \geq l_{\bullet}^*} \sum_{j \geq l_{\circ}^*} \frac{\binom{V_{\bullet}}{i} \binom{V_{\circ}}{j} \binom{V - (V_{\bullet} + V_{\circ})}{L - (i+j)}}{\binom{V}{L}}. \tag{7}$$

The quantity above is the multinomial version of the *surprise*, originally proposed to carry out a *community detection* exercise[51]. In our case, $L$ is the total number of links observed in our projections, while $V$ is the total number of possible links, i.e. $V = \frac{N(N-1)}{2}$. The quantities marked with $\bullet$ ($\circ$) refer to the corresponding core (periphery) quantities: for example, $l_{\bullet}^*$ is the number of observed links within the core, while $l_{\circ}^*$ is the number of observed links within the periphery. The presence of three binomial coefficients allows three different "species" of links to be accounted for: the binomial coefficient $\binom{V_{\bullet}}{i}$ enumerates the number of ways $i$ links can redistributed *within* the core, the binomial coefficient $\binom{V_{\circ}}{j}$ enumerates the number of ways $j$ links can redistributed *within* the periphery, and the binomial coefficient $\binom{V - (V_{\bullet} + V_{\circ})}{L - (i+j)}$ enumerates the number of ways the remaining $L - (i+j)$ links can be redistributed *between* the two, i.e. over the remaining $V - (V_{\bullet} + V_{\circ})$ node pairs (see Supplementary Note 3 for more details).

The mesoscale structure characterizing all discursive communities consists of a group of well-connected vertices linked to a group of low-degree, loosely inter-linked nodes, see Figs. 5, 6 and 7. Such a structure is known as *core-periphery* and is present in many social, economic and financial systems [52]. Remarkably, nodes belonging to the innermost shell overlap with the core ones computed with the multinomial version of the *surprise*, as proved by computing the *Jaccard index* over two sets of nodes which is a measure of similarity between two sets of elements and is defined as the size of the intersection divided by the size of the union of the two sets: $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$.

As a last comment, let us explicitly show the evolution of the number of nodes belonging to the core and to the periphery for each discursive community. As Fig. 8 shows, the core size is nearly constant throughout all the considered period while the periphery size rises in correspondence of the Election Day, showing a peak in correspondence of the day after the elections (i.e. 5 March 2018). This behavior, common to all communities, seems to indicate that, as the Election Day approaches, the number of topics discussed does increase.

**Filtering the projection.** Let us now focus on the structural features of the filtered projections. Before presenting the results of the analysis, we briefly recall how the filtering procedure works.

Filtering allows the detection of statistically significant o-occurences of hashtags measured on the initial bipartite network. More in detail, for each couple of hashtags, we count how many users are employing that specific couple. Then, we consider as a benchmark an ensemble of networks that, on average, preserves some information of the real user-hashtag bipartite network, i.e. the total number of links (as in a bipartite Erdös-Rényi, or Bipartite Random Graph, *BiRG*), the degree sequence of the hashtag layer (Bipartite Partial Configuration Model, *BiPCM*) or the degree sequence of both layers (Bipartite Configuration Model, *BiCM*). The larger the number of contraints (i.e., the properties preserved in the ensemble), the more detailed the description provided

by the corresponding benchmark and fewer links are validated. In this sense, BiCM is the most restrictive among the filtering null models while the BiRG is expected to retain the highest number of links between hashtags.

Such a procedure has been implemented in previous studies to detect the backbone of the network structure, filtering the real system from random noise, and to highlight non trivial behaviors in the original system[35,53]. In the present case, the filtering procedure allows us to distinguish *extremely popular hashtags* from these building proper *narratives*: the former, in fact, are just nodes with large degree, a feature that is compatible with at least one of the null models considered here and thus filtered out by our procedure; the latter, on the other hand, will likely be constituted by groups of hashtags whose non-trivial co-occurrence will survive the filtering procedure. As an example, let us think of a *single* popular message containing *several* hashtags used, however, only *once*—as it happens during the electoral campaign of some of the candidates that tweet the names of the towns visited during a specific day. In this case, the bipartite degree of the hashtags appearing *once*, but *together*, is given by the number of times the original message has been retweeted—plus the contribution of the original message. Hence, the number of times these hashtags co-occur basically coincides with their degree and the probability for their overlap to be validated becomes large (the nullmodel, in fact, would distribute the overlap among *all* hashtags). In other words, our filtering algorithm does not discard the hashtags appearing in viral messages which still contribute to the development of a narrative.

As mentioned above and as Fig. 9, 10 and 11 show, the overall effect of adopting a filtering procedure—regardless of its peculiarities—is that of reducing the total volume of the semantic networks. Differences exist, instead, when it comes to the analysis of nodes' mean degree. Particularly interesting in this respect is the behavior of the semantic networks of the M5S discursive community whose mean degree is affected to a much lesser extent by the BiRGM-induced filtering than those of the CDX and the CSX communities. This, in turn, implies that the information encoded into the total number of (re)tweets of the M5S bipartite user-hashtag network is able to account for the co-occurrences between any two hashtags less effectively than for the CDX and the CSX configurations. Equivalently, we may say that the structure of the M5S bipartite user-hashtag network requires less trivial information to be explained and the BiRGM (which is the simplest filter) recognizes it as significant.

For what concerns topics persistence, the ranking observed above with reference to the non-filtered projection basically coincides with the ranking observed on the filtered ones. Regarding topics centrality, instead, we observed that the filtering procedure with increasingly restrictive benchmarks makes "emerge" previously unscreened hashtags (e.g. *#sicurezza*, *#fallimentocinquestelle* and *#precariato*, respectively for the semantic networks induced by the CDX, CSX and M5S discursive communities). Centrality (e.g. the betweenness variant) is, in fact, a highly non-trivial feature that, generally speaking, is not reproduced by the information encoded into the degree sequence alone (not to mention the one encoded into the number of links): in fact, as Figs. 9, 10 and 11 clearly show, only the hashtags belonging to the innermost shells survive our filtering procedure.

Let us now move to discuss the mesoscale structure of the filtered projections looking, as usual, at one specific day that shows the richest structure (again 19 February 2018). Filtering the projections by adopting an increasingly restrictive benchmark make the projection sparser while letting less trivial structures emerge. Interestingly, the core portion of the semantic network corresponding to the M5S discursive community survives the most restrictive filtering (i.e. the BiCM-induced one), signalling the presence of a non-trivial group of keywords constituting the bulk of the communication in that community (see Fig. 9). Moreover, basically *all* hashtags representing topics of interest of the 2018 Italian electoral campaign persist. The same conclusion holds true for the number of triadic closures observed in correspondence of mediated events: their number is significantly larger with respect to a network model accounting for the total number of tweets only. This result is in line with what has been found for other socio-economic systems (e.g. the World Trade Web[54]) whose abundance of triadic closures is not reproduced by a benchmark model constraining the total number of links only.

In the following we will describe with more details the main characteristics of the filtered projections of the various semantic networks.

**The CDX discursive community.** Figure 9 depicts the semantic network of the center-right alliance on 19 February 2018.

In the BiCM projection, few links survive. In this situation, it becomes almost inappropriate to talk about communities, since we can find only links connecting two otherwise isolated nodes or small cliques. Nevertheless, even these few hashtags carry important information regarding the keywords used during the electoral campaign. This is the case of the cluster including *#stopislam*, *#stopinvasione* (*stop the invasion*), *#cdm* (i.e., the acronym for the Italian Council of Ministers) and *#forzeordine* (*law enforcement agencies*), asking for stronger countermeasures to the immigration flows from Northern Africa, perceived as dangerous for the security and for Italian cultural identity. On a similar topic, there is a clique composed by *#roma*, *#labaro* and *#primaporta*: the hashtags refer to neighborhoods in Rome, in which, during the days of the data collection, some thefts in apartments were reported. These hashtags were used to criticise the city administration of Rome, run by Virginia Raggi of the M5S. Moreover, a pair of nodes which represents insulting nicknames for the political opponents are connected between themselves. Those hashtags, *#pdioti* (i.e., an hashtag mixing the acronym *PD* and the word *idiots*) and *#m5stellisti*, are present in a popular message displaying both hashtags. *#m5stellisti*. There is also a clique formed by *#casapound* (i.e., a neo-fascist party), *#rai2* (i.e., the second channel of the national TV public service) and *#19febbraio*. This clique is the result of a viral tweet intended to advertise the presence of the leader of Casa Pound in a public debate held on *Rai2* on 19 February. Finally, the last clusters present in the BiCM-induced projection are more institutional: the first contains *#torniamoagovernare* (*let's go back to govern*), *#elezioniregionali2018* (*2018 regional administrative election*) and *#salvini*, while the other one is composed by *#flattax*, *#programma* (*program*) and *#veneto*. The latter set of hashtags is related to an event where the flat taxation is presented as part of the electoral program.
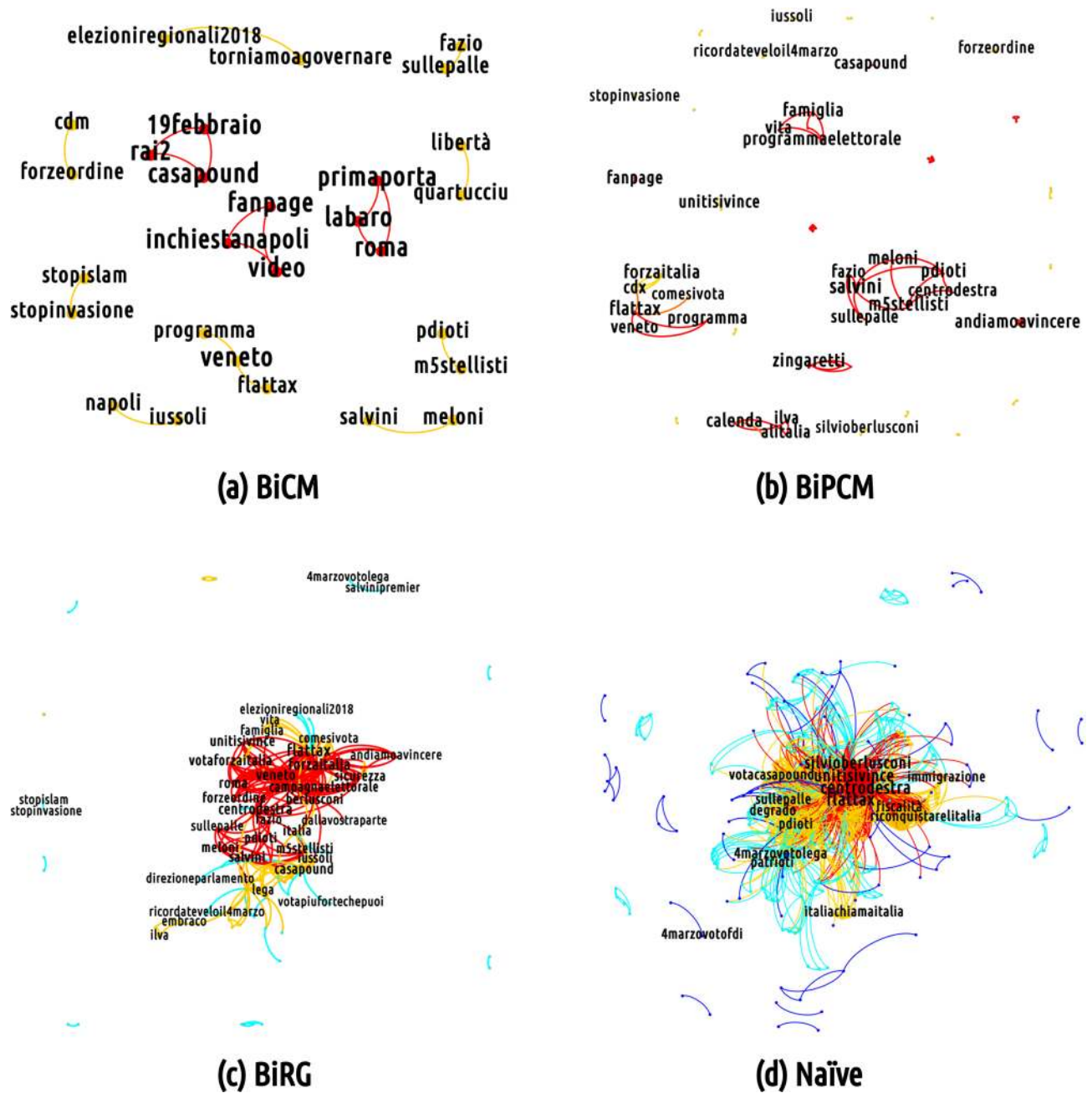
**Figure 9.** Mesoscale structure of (from bottom-right, clockwise) the non-filtered projection of the semantic network corresponding to the CDX discursive community on 19 February 2018 and of the projection of the same network filtered according to the BiRG, the BiCM and the BiPCM, respectively. The BiCM lets only *few* hashtags survive, reading *#iussoli*, *#sicurezza*, *#stopinvasione*, *#stopislam*. (Picture realized with Gephi software version 0.9.2. For more information about Gephi[49]: https://gephi.org/).

The BiPCM projection displays a structure in which the various sub-groups described above are reinforced (for instance, the set consisting of *#flattax*, *#programma* and *#veneto* is closed in a clique) and introduces new topics as *#calenda* (the Minister of the industrial development at the time of the electoral campaign) *#ilva* and *#alitalia*, respectively the greatest European steel factory which had severe problems of environmental, health and economic sustainability, and the Italian national airline, which has been at default risk in recent years. These hashtags are intended to criticize the action of the government in charge at that time. Interestingly, another cluster, related to the communication strategy of the most radical part of the center-right alliance, is detected by the validated BiPCM projection. The hashtags need a bit of context: during the electoral campaign, the journalist Fabio Fazio invited politicians from all political coalitions to his TV program *Che Tempo Che Fa* broadcast on the national television service, to promote their campaign. Fazio has been accused by all political forces of being too condescending with their opponents. Salvini refused Fazio's invitation, publicly stating his aversion for the journalist with insulting language. The hashtags *#sullepalle*, *#fazio*, *#salvini* together with other related to

(a) BiCM

(b) BiPCM

(c) BiRG

(d) Naïve
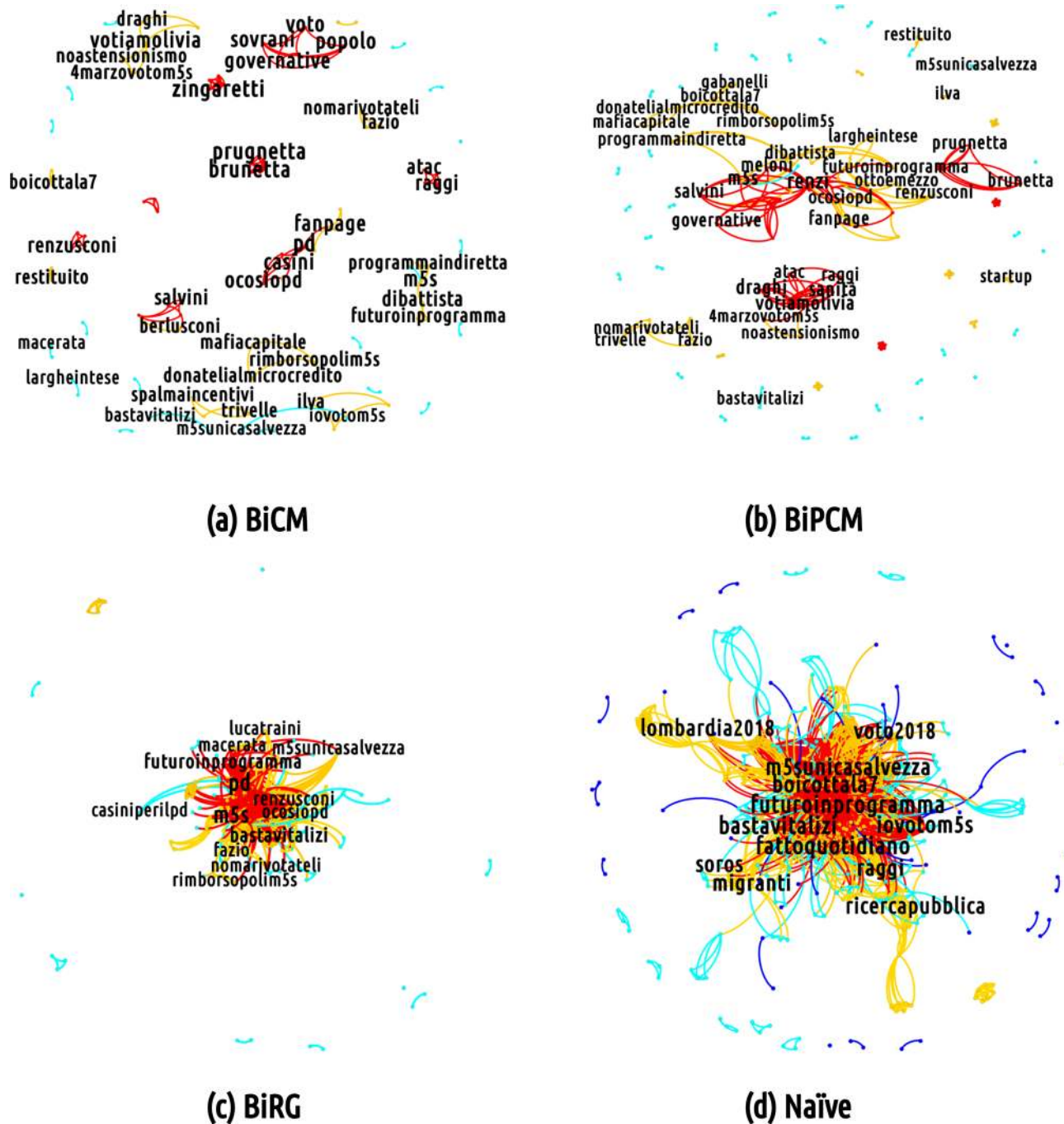
**Figure 10.** Mesoscale structure of (from bottom-right, clockwise) the non-filtered projection of the semantic network corresponding to the M5S discursive community on 19 February 2018 and of the projection of the same network filtered according to the BiRG, the BiCM and the BiPCM, respectively. The core portion of this network survives the most restrictive filtering (i.e. the BiCM-induced one), indicating that basically *all* hashtags representing topics of interest of the 2018 Italian electoral campaign persist. (Picture realized with Gephi software version 0.9.2. For more information about Gephi[49]: https://gephi.org/).

right-wing campaign topics such as #*vita* and #*famiglia* (*life* and *family*, related to the Italian anti-abortion movement) can be found in this cluster.

In the BiRG validated projection, the clusters found in the previous stricter projections are merged together to form a network organised along two poles: the first is more institutional with keywords related to the electoral campaign of 'Forza Italia', including hashtags such as #*campagnaelettorale* (*electoral campaign*), #*unitisivince* (*united we will win*), #*votaforzaitalia* (*vote Forza Italia*); the second is linked to the other two right-wing parties with both the names of their leaders (#*salvini* and #*meloni*), but also including their opponents, as #*pd*, #*renzi*, #*pdioti* and #*m5stellisti*. Interestingly, both poles are organised in a core-periphery: the two cores are connected by the hashtag #*centrodestra* (*center-right*), the peripheries by #*casapound* (the neo-fascist party cited above).
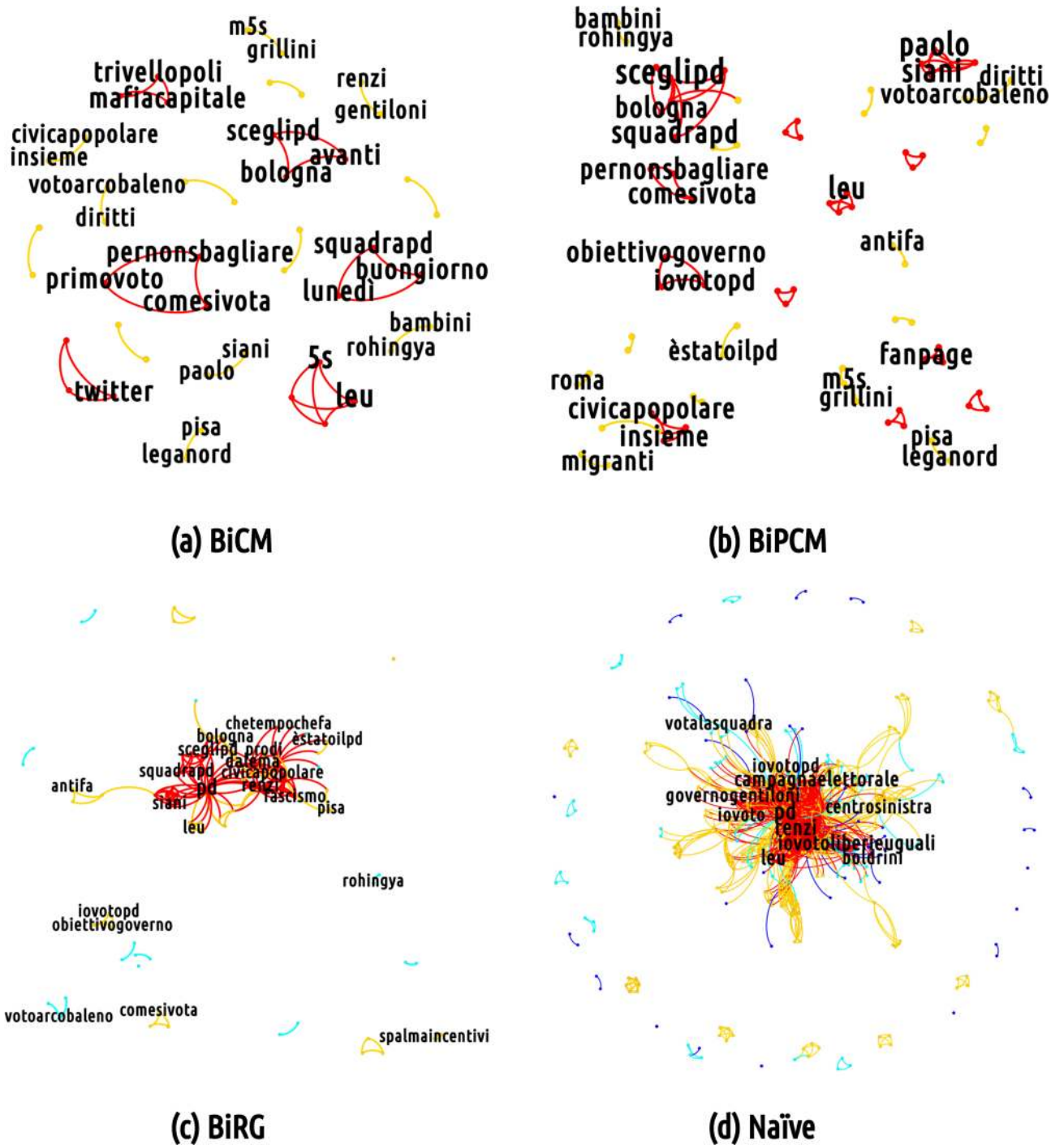
**Figure 11.** Mesoscale structure of (from bottom-right, clockwise) the non-filtered projection of the semantic network corresponding to the CSX discursive community on 19 February 2018 and of the projection of the same network filtered according to the BiRG, the BiCM and the BiPCM, respectively. The core portion of this network just partially survives the most restrictive filtering (i.e. the BiCM-induced one), while it is present in the less strict filtering (the BiPCM and the BiRG induced), representing a structure in between the stronge persistence of the M5S semantic network of Fig. 10 and the CDX one depicted in Fig. 9. (Picture realized with Gephi software version 0.9.2. For more information about Gephi[49]: https://gephi.org/).

**The M5S discursive community.** Overall, the communication strategy of the M5S is peculiar since its users tend to discuss a higher number of topics with respect to the other discursive communities. The evidence of a larger amount of validated hashtags can be explained by a higher number of retweeted messages that are able to generate peculiar hashtag co-occurences. In this sense, Twitter users in the M5S community appear to be more coordinated and thus manage to give their hashtags more visibility.

Interestingly, semantic networks corresponding to the M5S discursive community, shown in Fig. 10 , displays a rich structure. The core portion of the semantic network survives the BiCM filtering, signalling the presence of non-trivial group of keywords constituting the bulk of communication in that community. In particular, several clusters can be found, including the name of the opponents (*#renzi*, *#salvini*, *#gentiloni*, *#pd*), or a few nicknames assigned to them (as *#prugnetta*, *little plum*, for Brunetta, a member of 'Forza Italia'; *#renzusconi*, a mix between Renzi and Berlusconi, intending that there is little difference between the two of them) and other slogans taunting rivals (*#votiamolivia*, *let's vote them away*; *#nomarivotateli*, *no, but vote them again*, ironically targeting PD supporters; *#ocosìopd*, *this way or PD's way*, advertising the only political alternative to PD is the M5S). A few clusters represent political events in the electoral campaign. For instance, a cluster following the electoral campaign tour of Di Battista, a key member of the M5S, appears in this projection. Even a clique advertising a live streaming on Facebook can be observed, discussing the management of the public health system in the Lazio region governed by the PD, with the hashtags *#lazio*, *#sanità* (*public health system*), *#sancesareo* (i.e., the town were the live streaming was set), *#zingaretti* (i.e., Nicola Zingaretti who is the president of the Lazio region).

The topic of bad governance of political opponents represents a big part of the semantic network of the M5S community: in addition to the cluster mentioned above, another cluster focuses on the news about a journalist who was attacked during a campaign event held by the center-left coalition in Naples (*#fanpage*, which is the online newspaper for which the journalist worked; *#inchiestanapoli*, *Naples investigation*; *#video*). Moreover, a few hashtags refer to the scandal of a criminal organisation in Rome bribing members of established political parties. The 'Movimento 5 Stelle' expelled its representatives involved in this investigation and proposed that other parties do the same and transfer the amount of the bribe to the microcredit: *#donatelialmicrocredito* (*give them to the microcredit*), *#rimborsopoli* (*refund scandal*) all refer to this issue. There are also clusters targetting harsh debates. It is the case of *#dibiase*, referring to Letizia Di Biase who is the wife of the Italian Minister of Cultural Heritage and Activities. After being elected member of the council in the city of Rome, she did not resign when elected as member of the council of the region of Lazio. Di Biase was also criticised for blaming the major of Rome Virginia Raggi of the 'Movimento 5 Stelle', for filing for bankruptcy the city agency of public transportation while salvaging the regional one operated by the regional administration of the PD. Finally, there are traces of the debate between the virologist Roberto Burioni and the Head of the Italian Order of Biologists Vincenzo D'Anna concerning the presence of no-vax groups and M5S supporters during a national conference of the order of biologists. Other hashtags refer to Giorgia Meloni (i.e., the leader of 'Fratelli d'Italia') and the charges moved against her for her sympathy for neo-fascist parties and ideology.

The BiPCM projection increases the connections among the topics and a few hashtags appear to be related to names of places covered by the campaign tour of Carlo Sibilia, another member of the M5S. Instead, the BiRG projection displays a strong core-periphery structure.

**The CSX discursive community.** In the case of the CSX community, the number of hashtags used is relatively small. Compared to other discursive communities, the semantic network of the center-left group validated by the BiCM (see Fig. 11) focuses more on political topics, as proven by the pair *#diritti* (*rights*) and *#arcobaleno* which refer to the LGBTQIA+ civil rights, but also by the coupling of *#bimbi* (*children*) and *#rohingya* both related to the topic of the Rohingya exodus in Myanmar and the condition of children. Other clusters pivot around instructions for youngsters voting for the first time (*#primovoto*, *first vote*; *#comesivota*, *how to vote*; *#pernonsbagliare*, *how to avoid mistakes*) and call for fact checking during the electoral campaign, with hashtags *#factchecking* and *#checkpolitiche2018*.

Interestingly, a clique is formed by hashtags of a single popular tweet *#trivellopoli*, *#mafiacapitale* and *#consip*, i.e. three scandals in which the PD was involved. This tweet suggested that those scandals suspiciously appeared during the electoral campaign in order to damage the name of the 'Partito Democratico' thereby limiting its performance at the general elections.

Another conversational line unfolds along the candidacy of Paolo Siani, a physician particularly active in providing support, in collaboration with local NGOs, to children of the poor neighborhoods of Naples at risk of being recruited for organized criminal activities. More broadly, the public presentation of the Italian 'Partito Democratico' candidates team constitutes a frequent argument debated in the community as shown by the two hashtags *#renzi* and *#gentiloni*, respectively the PD secretary and the PD prime minister during the electoral campaign, and in the cliques *#bologna*, *#avanti* (*let's move forward*) and *#sceglipd* (*choose PD*); and *#lunedì* (*Monday*), *#buongiorno* (*good morning*) and *#squadrapd* (*PD team*). The former refers to an event led by the secretary and Prime Minister candidate Matteo Renzi in Bologna, the latter appeared in a message promoting a massive electoral campaign, due to the probable uncertainty of the result of the election.

In the BiPCM validated projection more connections appear, enriching topics which had already emerged, as in the case of the candidacy of Paolo Siani mentioned above: *#babygang*, *#napoli* (*Naples*), *#infanzia* (*childhood*) merge with the previous hashtags *#paolo* and *#siani*. A new cluster containing the name of the opponents (*#dimaio*, *#salvini*, *#meloni*, *#fascismo*) is also present.

In the BiRG validated projection, the aforementioned structures gain new links and new nodes and a richer structure becomes evident. In particular, three main communities appear: the first (in orange in Fig. 11) pivoting around political adversaries (including *#salvini*, *#meloni*, *#dimaio*, *#grillini*, *#berlusconi*); the second advertising political subjects and events of the electoral campaign (including *#sceglipd*, *choose PD*; *#squadrapd*, *PD team*; *#diritti*, *rights*, and so on); and the third concerning to the candidacy of Paolo Siani. A peripheral clique can be found with hashtags advertising the event in Venice of *Liberi e Uguali*, a political party on the left of PD (*#antifa*, *#liberieuguali*, *#venezia*, *Venice*).

*Final remarks on the filtering procedure.* In all naïve projections we observe a rich structure, with a particularly evident core-periphery organisation. This structure is progressively disintegrated through filtering in various ways and depending on the strictness of the benchmark used. While this disintegration affects semantic structures generated by all discursive communities, the various groups display a different resilience to the filtering procedure, the one revolving around M5S acoounts being that with the least trivial structure, hence being affected the least. Quite relevantly, the different network structures carry information about the strategy followed by the various discursive communities during their political campaign.

The validation procedure proposed in[35] projects non-trivial co-occurrences of links in the bipartite networks, i.e. those that are not explained by the ingredients of the null-model used for filtering. In this sense, validated nodes in the projection are not necessarily those with, for example, the highest (bipartite) degree, but those grappling to other hashtags in the semantic network to a larger extent than expected by just looking at the original bipartite network. With respect to the examined case study, the validated projections show that the more the validated links, the more hashtags are used to refer to a single subject, against the random superposition of ubiquitous slogans. This seems to be the case particularly for M5S community while it is true to a much lesser extent for the CDX discursive community where the amount of nodes in the BiCM validated projection is extremely limited. The validation procedure allows us to focus on the least trivial connections and thus to observe different conversational lines that shape the political communication of the various discursive communities. In this way, the validation procedure allows to uncover otherwise invisible conversational lines that shape the Twitter activity of each discursive community.

In the CDX, a clear thematic distance is present between the far right (formed by partied led by Matteo Salvini and Giorgia Meloni) and center-right politicians (Silvio Berlusconi and his party 'Forza Italia') in terms of topics and electoral slogans. While the former insists on security issues related to migration flows from Northern Africa, the latter tends to promote a united center-right alliance. There is an evident semantic diversification with completely different keywords used in the tweets: the far right uses more aggressive statements and bad words, while the second is more reassuring and institutional.

The M5S projected semantic networks are especially rich in structure, due to the strong usage of hashtags in this community. Most of them are referring to political opponents with nicknames and ironic slogans. A great part of the filtered semantic network is devoted to highlight the deceitfulness of the M5S opponents.

Validated semantic networks of the CSX community are poorer than those of M5S, but richer than those of the CDX. Their major feature is to present mostly events of the electoral campaign, their candidates at a national and regional level and the weaknesses of their political opponents.

It is worth noticing that the peculiarities of the three discursive filtered semantic networks are present in other days which are not explicitly commented here. For instance, on 11 February, we can still observe two different poles of the debates in the CDX community, one promoted by the supporters of 'Forza Italia' and the other promoted by the supporters of far-right wing parties. As observed for the 19 of February, the two poles use different vocabularies and focus, respectively, on reforming taxation and labour or on migration issues. Analogously, the M5S displays a cluster of discussion against the use of vaccines, few clusters against an alleged *quid pro quo* between some PD members and some businessmen as well as some other cluster teasing political opponents. Finally, the CSX focus on the election candidates presentation and few national problems (increasing inequality, poverty and the decreasing birth-rate). In all three networks there are also mentions to the demonstration involving nearly thirty-thousand persons against neo-fascism held in Macerata on 10 February, albeit with different levels of attention.

## Discussion

In this section, we would like to discuss more in detail some relevant features of our methodological approach and to summarize results about the Italian case we examined throughout this work. Our study focused on the activity on Twitter of three different discursive communities tied to the main political topics and actors competing during general elections in Italy in 2018. It exploited approximately 1 million of tweets which we used to define networks of statistically significant co-occurrences of hashtags at a daily time scale. Compared to extant research, we argue that our proposed approach innovates studies in this domain in two main ways. First, we propose a methodological framework[11,44] to investigate the semantic components of online political discussions coupling attention for topics prominence and persistence with a sound exploration of how they are organized and structured along conversational lines. Second, we did not treat the semantic aspect in isolation but, in fact, built a tight connection between contents discussed by users and the type of interactions they performed via Twitter. Moreover, our way of isolating discursive communities and related semantic networks does not entail any manual intervention. Initial distinction between verified and non-verified accounts is not defined by researchers but is, in fact, assigned by the Twitter platform itself. Relevantly, as the categorization of accounts according to these two categories is regularly provided through the Streaming API by the platform itself, the solidity of the initial partition is ensured platform-wise. Similarly, topics and conversational lines within semantic networks are not biased by construction but, in fact, are a direct output of the application of filtering procedures with maximally random benchmarks.

Against this background, we sought to offer a solid analysis of the public discussion developed online during a crucial political event and to infer, starting from semantic aspects, marking traits of the different political identities that guide political courses of action in the Italian political scenario. One of the main findings of this paper concerns the way the topological structure of semantic networks "reacts" to the so-called mediated events (i.e. TV debates, the media coverage of offline events, etc.) thus revealing not only different sensibility towards the media sphere but, more significantly, different identity traits that characterize each discursive political community. The topology of the CDX community is strongly dependent on these events (the mean degree of nodes

increases in correspondence of specific TV shows), meaning that this group of users is more involved in the activity of retweeting during the appearance of political actors particularly on television. Conversely, the activity of the M5S community appears to be much more "distributed". In fact, although M5S supporters are sensitive to TV shows as well, their retweeting activity is not exclusively driven by media events but, rather, follows their preference for a generalized use of social media for organizational and political communication activities. Finally, the activity of the CSX community is characterized by a somewhat "intermediate" behavior: even when mediated events affect the Twitter discussion, the attention of the whole community is somehow "shared" among the various actors constituting the center-left alliance.

Particularly insightful is the analysis of our semantic networks at the mesoscale: what emerges is the presence of a core of topics, i.e. a densely-connected bulk of hashtags surrounded by a periphery of loosely interconnected (sub-)topics. This indicates that daily semantic networks are characterized by few relevant hashtags to which other, less relevant ones, attach. This structure is maintained during the whole observation period and differences emerge only with respect to the number of *peripheral* themes entering the discussion. The resilience of the core-periphery structure is not the same for the various discursive communities. In the context of semantic networks, the fact that the system is more or less resilient to the filtering implies that the various political groups have developed differently their political narrative, focusing their communications on few related terms per subject or mentioning a set of omnipresent hashtags in all messages. Even in the response to the filtering procedure, the M5S and the CDX communities represent the two extremes, displaying respectively the most and the least resilient semantic network.

These differences are the effect of various styles used in writing posts. When targeting a specific theme, members of the M5S community use several hashtags that are subsequently re-used by other users writing on the same argument in order to make keywords and slogans more recognizable and visible. Conversely, in the CDX community, the number of hashtags per message is more limited and users tend to use particularly viral hashtags. Moreover, the CDX shows a diversified communication strategy, hence a possible internal political fracture, due to the different approaches of the various parties in the alliance: right-wing politicians are more aggressive towards opponents, while center-right ones tend to focus on unitive (for the coalition) keywords. Somewhere in between, the CSX community balances a use of hashtags to criticize its adversaries but also to amplify the reach of events and initiatives of the electoral campaign.

## Conclusions

Social media platforms have dramatically changed patterns of news consumption and, over the last years, they have become increasingly central during political events, especially electoral campaigns. In this respect, Twitter has been shown to play a major role, thus attracting the attention of scientists from all disciplines. So far, however, researchers have mainly focused on users' activity, paying little attention to semantic aspects which are instead particularly relevant to detect online debates, understand their evolution and, ultimately, inferring the behavioral rules driving (online but also offline) electoral campaigns.

As claimed in the section "Introduction", our goal in this paper is to advance a comprehensive methodological framework to fill a persistent attention gap for the semantic aspects of online political discussions. Accordingly, we did put forward a scheme of analysis that couples attention for more evident aspects, such as topic visibility, persistence and strategic uses in conversation, with the systematic analysis of more invisible, and yet crucial aspects of content production and circulation—particularly, the identification of conversational cores and main lines of development. In doing so, we believe that our proposed approach provides a solid starting point to understand the symbolic aspects that flank and, in fact, nurture current dynamics of online civic participation, political partisanship and polarization, which have so far catalyzed attention within the research community.

More relevantly, our proposed approach did not bring prominence to semantic aspects leaving the social side of online dynamics behind. The semantic networks we analyzed in this paper follows from an identification of discursive communities that leans on an entropy-based framework, which is a methodological advancement per se[11,44]. On the one hand, this technique allowed us to filter the retweeting activity of users while singling out the statistically-relevant information at the desired level of detail. Consistently, the semantic structures we induce from communities are indicative of users' political affiliation yet without passing through any manual labelling of our media contents (e.g. as performed in[32,34]). On the other hand, we deduce the discussions taking place in each community by connecting any two hashtags if used a signicantly large number of times by the users of that community, hence overcoming the limitations of other analyses[26,29] where online political conversations are studied in an unrelated fashion with respect to the relational system amongst users sustaining them.

Our proposed approach allowed to reach several fine-grained insights about the Italian case both at the macro level, grasping the semantic peculiarities of broader conversations taking place within discursive communities, and, at the micro level, narrowing down our exploration to single and meaningful points in time during the electoral campaign period. Above and beyond the particular case study we analysed, we believe that the same approach can be applied to disentangle the intricacies of large-scale Twitter conversations in all domains and regardless of the language in which they take place with the only requirement of extracting the data set with a meaningful list of anchor hashtags.

Paralleling these advantages, our approach remains limited mainly in two respects. First however detailed and multilevel, our semantic analysis remains a partial investigation of the symbolic universe that is produced and circulated online in conjunction with relevant political events and dynamics. The semantic structures we investigate in this paper are formed in the space created by a single platform and pivot around the use of a specific feature for marking contents—i.e., hashtags. Ad-hoc publics that assemble around topics, in fact, are not exhausted by communities that form on particular social media platforms—let alone around specific hashtags[55]. Moreover, looking at conversations forming around specific hashtags fails to include those contributions that,

albeit pertinent, are delivered without including any specific content markers[56]. The very procedure to discount all those messages that did not contain an hashtag from the initial corpus of tweets containing election-related keywords well illustrates the need to account for multiple communicative actions performed by users.

Second, albeit automatically inferred and not depending in any way from aprioristic assumptions on users' political orientation, our discursive communities are neither representative of the Italian voting population nor an exact indication of users' political affiliation. On the one side, it is widely recognized that Twitter users are not representative of broader populations (in this case, of all Italian citizens) and that Twitter Search API does not guarantee the representativeness of the data themselves[57]. On the other, the identification of discursive communities starts from retweets which, as mentioned above, simply express an explicit bestowal of attention. The total lack of information about the actual political afliation of users does severely limit our capacity to predict, starting from semantic data, relevant political outcomes such as voters turnout, election results, or the construction of political alliances between parties.

Nonetheless, without any claim of exhaustivity, our mapping of the Twitter discussion in occasion of national elections provides a useful entry point to reason around the online construction of political collective identities. A plethora of studies based on Twitter freely available data has shown that it is indeed possible to infer the political orientation of users from tweets and to analyze electoral debates and societal discussions[11–13,32,44,58] shedding light on the political implications of non-traditional political acts such as expressing publicly on social media. Users employing in their tweets election-related keywords and hashtags did in fact contribute to frame the electoral campaign period along certain lines and they did so upon a platform that was not only widely diffused in Italy in that specific moment (According to Audiweb[59], 9 millions of Italian users were active on Twitter in 2018) but that also plays a pivotal political communication role[40]. Moreover, the strict filtering procedure we applied leads to the identification of networks with only statistically-relevant hashtags information, as the projections that we adopted guarantees that the analysis of induced semantic networks is sound from both a methodological and interpretative point of view. Thus, albeit non representative of and non generalizable to the overall Italian population, both discursive communities and induced semantic networks that we examined can be thought as a solid starting ground to develop more fine-grained studies of voters' political opinion and behaviors.

## References

1. Eurobarometer. *Standard Eurobarometer 92 "Media Use in the European Union" Report. European Commission—Public Opinion*. https://doi.org/10.2775/80086. Accessed 10 Mar 2021 (2019).
2. Matsa, K. E. & Shearer, E. *News Use Across Social Media Platforms 2017. Pew Research Center*. Accessed 30 Sep 2018 (2018).
3. Schmidt, A. L. *et al.* Anatomy of news consumption on Facebook. *Proc. Natl. Acad. Sci.* **114**, 3035–3039. https://doi.org/10.1073/pnas.1617052114. http://www.pnas.org/content/114/12/3035.full.pdf (2017).
4. Pavan, E. The integrative power of online collective action networks beyond protest. exploring social media use in the process of institutionalization. *Soc. Mov. Stud.* **16**, 433–446. https://doi.org/10.1080/14742837.2016.1268956 (2017).
5. Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E. & Quattrociocchi, W. Modeling confirmation bias and polarization. *Sci. Rep.* **7**, 40391. https://doi.org/10.1038/srep40391. arXiv:1607.00022 (2017).
6. Schmidt, A. L., Zollo, F., Scala, A., Betsch, C. & Quattrociocchi, W. Polarization of the vaccination debate on facebook. *Vaccine* **36**, 3606–3612. https://doi.org/10.1016/j.vaccine.2018.05.040 (2018).
7. Zollo, F. *et al.* Debunking in a world of tribes. *PLOS ONE* **12**, 1–27. https://doi.org/10.1371/journal.pone.0181821 (2017).
8. Morales, A. J., Borondo, J., Losada, J. C. & Benito, R. M. Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos Interdiscip. J. Nonlinear Sci.* **25**, 033114. https://doi.org/10.1063/1.4913758 (2015).
9. Cherepnalkoski, D. & Mozetič, I. Retweet networks of the European parliament: Evaluation of the community structure. *Appl. Netw. Sci.* **1**, 2. https://doi.org/10.1007/s41109-016-0001-4 (2016).
10. Pei, S., Muchnik, L., Andrade, J. S., Jr., Zheng, Z. & Makse, H. A. Searching for superspreaders of information in real-world social media. *Sci. Rep.* **4**, 5547. https://doi.org/10.1038/srep05547. arXiv:1405.1790 (2014).
11. Becatti, C., Caldarelli, G., Lambiotte, R. & Saracco, F. Extracting significant signal of news consumption from social networks: The case of Twitter in Italian political elections. *Palgrave Commun.* https://doi.org/10.1057/s41599-019-0300-3. arXiv:1901.07933 (2019).
12. Caldarelli, G., De Nicola, R., Del Vigna, F., Petrocchi, M. & Saracco, F. The role of bot squads in the political propaganda on Twitter. *Commun. Phys.* **3**, 1–15. https://doi.org/10.1038/s42005-020-0340-4. arXiv:1905.12687 (2020).
13. González-Bailón, S., Borge-Holthoefer, J. & Moreno, Y. Broadcasters and hidden influentials in online protest diffusion. *Am. Behav. Sci.* **57**, 943–965. https://doi.org/10.1177/0002764213479371 (2013).
14. Castells, M. Network theory | a network theory of power. *Int. J.Commun.* **5** (2011).
15. Padovani, C. & Pavan, E. Global governance and icts: Exploring online governance networks around gender and media. *Glob. Netw.* **16**, 350–371. https://doi.org/10.1111/glob.12119 (2016).
16. Bastos, M. T. & Mercea, D. Serial activists: Political twitter beyond influentials and the twittertariat. *New Med. Soc.* **18**, 2359–2378. https://doi.org/10.1177/1461444815584764 (2016).
17. Pariser, E. *The Filter Bubble: What the Internet is Hiding from You* (Penguin Group, 2011).
18. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A. & Tesconi, M. Fame for sale: Efficient detection of fake Twitter followers. *Decis. Support Syst.* https://doi.org/10.1016/j.dss.2015.09.003. arXiv:1509.04098 (2015).
19. Ferrara, E., Varol, O., Davis, C., Menczer, F. & Flammini, A. The rise of social bots. *Commun. ACM* **59**, 96–104. https://doi.org/10.1145/2818717 (2016).
20. Cresci, S., Petrocchi, M., Spognardi, A. & Tognazzi, S. From reaction to proaction: Unexplored ways to the detection of evolving spambots. In *Web Conference 2018—Companion World Wide Web Conference WWW 2018*. https://doi.org/10.1145/3184558.3191595 (2018).
21. Cresci, S., Petrocchi, M., Spognardi, A. & Tognazzi, S. On the capability of evolved spambots to evade detection via genetic engineering. *Online Soc. Netw. Med.*. https://doi.org/10.1016/j.osnem.2018.10.005 (2019).
22. Latour, B. *Reassembling the Social: An Introduction to the Actor-Network Theory* (Oxford University Press, 2005).
23. Gillespie, T., Boczkowski, P. J. & Foot, K. A. *Materiality and Media in Communication and Technology Studies: An Unfinished Project, 21–51* (The MIT Press, 2013).

24. Keller, T. R. & Klinger, U. Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Polit. Commun.* **36**, 171–189. https://doi.org/10.1080/10584609.2018.1526238 (2019).
25. Keller, T. R., Hase, V., Thaker, J., Mahl, D. & Schäfer, M. S. News media coverage of climate change in India 1997–2016: Using automated content analysis to assess themes and topics. *Environ. Commun.* **14**, 219–235. https://doi.org/10.1080/17524032.2019.1643383 (2020).
26. Evans, H. K., Smith, S., Gonzales, A. & Strouse, K. Mudslinging on twitter during the 2014 election. *Social Media + Society* **3**. https://doi.org/10.1177/2056305117704408 (2017).
27. Gaumont, N., Panahi, M. & Chavalarias, D. Reconstruction of the socio-semantic dynamics of political activist Twitter networks-method and application to the 2017 French presidential election. *PLOS ONE* **13**, 1–38. https://doi.org/10.1371/journal.pone.0201879 (2018).
28. Giglietto, F. *et al.* Mapping italian news media political coverage in the lead-up of 2018 general election. *SSRN Electron. J.* https://doi.org/10.2139/ssrn.3179930 *(2018)*.
29. Xiong, Y., Cho, M. & Boatwright, B. Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of twitter during the #metoo movement. *Public Relat. Rev.* **45**, 10–23. https://doi.org/10.1016/j.pubrev.2018.10.014 (2019).
30. Celli, F. & Rossi, L. Long chains or stable communities? The role of emotional stability in twitter conversations. *Comput. Intell.* **31**, 184–200 (2015).
31. Giglietto, F. & Lee, Y. To be or not to be charlie: Twitter hashtags as a discourse and counter-discourse in the aftermath of the 2015 charlie hebdo shooting in France. In *Proceedings of the 5th Workshop on Making Sense of Microposts Co-located with the 24th International World Wide Web Conference* 33–37 (2015).
32. Conover, M. D., Goncalves, B., Ratkiewicz, J., Flammini, A. & Menczer, F. Predicting the political alignment of twitter users. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing* 192–199. https://doi.org/10.1109/PASSAT/SocialCom.2011.34 (2011).
33. Hanna, A. *et al.* Partisan alignments and political polarization online: A computational approach to understanding the French and US presidential elections. In *Proceedings of the 2nd Workshop on Politics, Elections and Data*, PLEAD '13, 15–22. https://doi.org/10.1145/2508436.2508438 (Association for Computing Machinery, 2013).
34. Badawy, A., Ferrara, E. & Lerman, K. Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign. In *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '18, 258–265 (IEEE Press, 2018).
35. Saracco, F. *et al.* Inferring monopartite projections of bipartite networks: an entropy-based approach. *New J.Phys.* **19**, 053022. https://doi.org/10.1088/1367-2630/aa6b38 (2017).
36. Bobba, G. *et al.* Who's the winner? an analysis of the 2018 Italian general election. *Ital. Polit.Sci.* **1** (2018).
37. Bovet, A. & Makse, H. A. Influence of fake news in twitter during the 2016 US presidential election. *Nat. Commun.* **10**. https://doi.org/10.1038/s41467-018-07761-2 (2019).
38. Vicario, M. D., Zollo, F., Caldarelli, G., Scala, A. & Quattrociocchi, W. Mapping social dynamics on facebook: The brexit debate. *Soc. Netw.* **50**, 6–16. https://doi.org/10.1016/j.socnet.2017.02.002 (2017).
39. Jungherr, A. Twitter use in election campaigns: A systematic literature review. *J. Inf. Technol. Polit.* **13**, 72–91. https://doi.org/10.1080/19331681.2015.1132401 (2016).
40. Marchetti, R. & Ceccobelli, D. Twitter and television in a hybrid media system. *J. Pract.* **10**, 626–644. https://doi.org/10.1080/17512786.2015.1040051 (2016).
41. Small, T. A. What the hashtag?. *Inf. Commun. Soc.* **14**, 872–895. https://doi.org/10.1080/1369118X.2011.554572 (2011).
42. Hong, L., Convertino, G. & Chi, E. *Language Matters in Twitter: A Large Scale Study* (2011).
43. Gomaa, W. H. & Fahmy, A. A. A survey of text similarity approaches. *Int. J. Comput. Appl.* **68**, 13–18 (2013) **(full text available)**.
44. Radicioni, T., Squartini, T., Pavan, E. & Saracco, F. Networked partisanship and framing: a socio-semantic network analysis of the Italian debate on migration. *arXiv e-prints* arXiv:2103.04653 *(2021)*.
45. Lin, Y., Keegan, B., Margolin, D. & Lazer, D. Rising tides or rising stars?: Dynamics of shared attention on twitter during media events. *PLOS ONE* **9**, 1–12. https://doi.org/10.1371/journal.pone.0094093 (2014).
46. Bianconi, G., Darst, R. K., Iacovacci, J. & Fortunato, S. Triadic closure as a basic generating mechanism of communities in complex networks. *Physical Review E* **90**, 042806. https://doi.org/10.1103/PhysRevE.90.042806. arXiv:1407.1664 (2014).
47. Easley, D. & Kleinberg, J. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge University Press, 2010).
48. Kong, Y., Shi, G., Wu, R. & Zhang, Y. k-core: Theories and applications. *Phys. Rep.* **832**, 1. https://doi.org/10.1016/j.physrep.2019.10.004 (2019).
49. Bastian, M., Heymann, S., & Jacomy, M. *An Open Source Software for Exploring and Manipulating Networks* (Gephi, 2009).
50. Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893. https://doi.org/10.1038/nphys1746 (2010).
51. de Jeude, J. v. L., Caldarelli, G. & Squartini, T. Detecting core-periphery structures by surprise. *Europhys. Lett.* **125**, 68001 (2019).
52. Malvestio, I., Cardillo, A. & Masuda, N. Interplay between $k$-core and community structure in complex networks. 2005.01147 (2020).
53. Straka, M., Caldarelli, G. & Saracco, F. Grand canonical validation of the bipartite international trade network. *Phys. Rev. E* **96**. https://doi.org/10.1103/PhysRevE.96.022306. arXiv:1703.04090v1 (2017).
54. Saracco, F., di Clemente, R., Gabrielli, A. & Squartini, T. Randomizing bipartite networks: The case of the World Trade Web. *Sci. Rep.* **5**, 10595. https://doi.org/10.1038/srep10595. arXiv:1503.05098 (2015).
55. Highfield, T. & Bruns, A. Compulsory voting, encouraged tweeting? Australian elections and social media. In *The Routledge Companion to Social Media and Politics* (Enli, G., Bruns, A., Larsson, A. O., Skogerbo, E. & Christensen, C. eds.) 338–350 (Routledge, 2016).
56. Hanna, A. Computer-aided content analysis of digitally enabled movements. *Mobilization Int. Q.* **18**, 367–388 (2013).
57. Pfeffer, J., Mayer, K. & Morstatter, F. Tampering with twitter's sample api. *EPJ Data Sci.* **7**, 50 (2018).
58. A. Rapini, E. P. Antifascism retweeted. semantic networks around #antifa* in the italian twittersphere. In *Annual Conference of the Association of Italian Political Communication (AssoComPol)* (2019).
59. Cosenza, V. *Social Media in Italia: Utenti e Tempo di Utilizzo nel 2018*. Accessed 10 Mar 2021 (2019).

## Acknowledgements

### Author contributions

T.R., E.P., T.S. and F.S. contributed equally to outline the research, interpret the results, write and review the manuscript. T.R., T.S. and F.S. provided the analytical and numerical tools and T.R. performed the analysis.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-92337-2.

**Correspondence** and requests for materials should be addressed to T.R.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.