O **EPJ Data Science**
a SpringerOpen Journal

REGULAR ARTICLE   **Open Access**

Check for updates

Italian Twitter semantic network during the Covid-19 epidemic

Mattia Mattei[1,3*] (iD), Guido Caldarelli[2], Tiziano Squartini[3] and Fabio Saracco[3,4]

*Correspondence:
mattia.mattei14@libero.it
[1] University of Salento, P.zza Tancredi 7, 73100 Lecce, Italy
[3] IMT School for Advanced Studies, P.zza S. Ponziano 6, 55100 Lucca, Italy
Full list of author information is available at the end of the article

Abstract

The Covid-19 pandemic has had a deep impact on the lives of the entire world population, inducing a participated societal debate. As in other contexts, the debate has been the subject of several d/misinformation campaigns; in a quite unprecedented fashion, however, the presence of false information has seriously put at risk the public health. In this sense, detecting the presence of malicious narratives and identifying the kinds of users that are more prone to spread them represent the first step to limit the persistence of the former ones. In the present paper we analyse the semantic network observed on Twitter during the first Italian lockdown (induced by the hashtags contained in approximately 1.5 millions tweets published between the 23rd of March 2020 and the 23rd of April 2020) and study the extent to which various discursive communities are exposed to d/misinformation arguments. As observed in other studies, the recovered discursive communities largely overlap with traditional political parties, even if the debated topics concern different facets of the management of the pandemic. Although the themes directly related to d/misinformation are a minority of those discussed within our semantic networks, their popularity is unevenly distributed among the various discursive communities.

Keywords: Covid-19 epidemic; Twitter; Complex networks; Semantic networks; Disinformation; Misinformation

1 Introduction

The Covid-19 pandemic has had a deep impact on nearly every human activity; as such, it is not surprising that it has generated a widespread online debate which, in turn, attracted the interest of scholars from different disciplines [1–7]. Due to its unmediated nature, the online debate was affected by quite an amount of low-quality contents that, more than in other circumstances, had the potential to severely put at risk the public health. In this sense, the High Representative of the Union for Foreign Affairs and Security Policy explicitly exposed her concerns about the possible effects of a wrong communication, on social media, during the pandemic: "Disinformation can have severe consequences: it can lead people to ignore official health advice and engage in risky behaviour, or have a negative impact on our democratic institutions, societies, as well as on our economic and financial situation." [8].

Springer

Indeed, the already prolific research on the diffusion of d/misinformation [9–18] was enriched by the reactive attention of scholars. For instance, Gallotti et al. [3] released a real-time dashboard to monitor the risk of exposure to d/misinformation in the various countries on Twitter and proposed an Infodemic Risk Index, based on epidemic studies. In particular, they were able to detect early-warning signals related to the diffusion of d/misinformation campaigns. Interestingly, Celestini et al. [2] analysed the Italian debate on Facebook and found that controversial topics associated to known false information represented a limited amount of the traffic and, moreover, are less able to engage the audience than reliable media. Even more remarkably, authors observed the presence of a sort of small-world effect in the exposure to URLs, such that nearly any user can access any kind of news source, even if the navigation is limited to a reduced number of pages. Following a similar line of research, Yang et al. [5] analysed both Twitter and Facebook, focusing their attention on links to low-credibility contents; in particular, they observed the presence of a limited number of extremely influential accounts, i.e. the d/misinformation *super-spreaders*.

The comparison of several platforms is also the research target of [4]: Gab, Facebook, Instagram, Reddit and Youtube were analysed and, by fitting the information spreading via an epidemic model, an R_0 parameter was assigned to any platform. Even if the spreading patterns are similar, the various online social networks are differently exposed to the risk of d/misinformation.

Across the entire literature overview, a limited attention is paid to identify the narratives shared by the various users. In this sense, the contribution of [19, 20] is twofold: first, the authors attempt at inferring the various communication strategies on Twitter by leveraging on the different usage of the hashtags made by the users (and considering both the attitude of the various accounts towards the use of different hashtags and the popularity of the latter ones); second, they represent one of the few examples of applications of the entropy-based framework [21, 22] for the analysis of Online Social Networks. So far, in fact, entropy-based null-models have been successfully employed to analyse economic [23–27] and financial networks [28, 29], either to reconstruct such systems from limited information [30] or as a benchmark for the analysis of their network structure. Only recently, these methods have been applied to the analysis of Online Social Networks, evidencing non-trivial phenomena such as customer tastes in online retail networks [31], the presence of discursive communities on Twitter [17], the presence of coordinated activities of automated accounts [18], the extent to which the various discursive communities are exposed to d/misinformation campaigns [6].

In the present manuscript, we examine the Italian online debate during the peak of the first wave of the Covid-19 pandemic, following the approach of [19, 20], i.e. focusing our attention on the semantic side. To this aim, we represent the Twitter online debate as a bipartite network in which the two layers respectively represent accounts and hashtags and a link connect two nodes if the considered account has used at least once the given hashtag during the observation period. Then, we extract the semantic network of hashtags by employing an entropy-based null-model as a benchmark to obtain a monopartite representation of our system out of its original, bipartite structure [32]. The popularity of this formalism is due to its many desirable features: first, it allows one to identify the probability distribution which is maximally non-committal with respect to the missing information, i.e. the least-biased one by unjustified assumptions—as the ones *not* encoded

into the constraints [33]; second, it allows, either exact or approximate, statistical tests to be carried out via the proper definition of null hypotheses [34, 35]; third, its flexibility allows one to explore the effectiveness of a wide range of different constraints to either reconstruct [36] or detect [22] patterns—even outside the network realm [37–41].

Differently from [19, 20], where the focus was on the communication strategies of the different discursive communities (respectively, during the 2018 Italian electoral campaign and about the debate on migration policies), here we produce a single semantic network across the entire period of the debate. The various communities engage in different discussions in different moments, depending on the interest towards the topic. Nevertheless, as also observed by focusing on URLs news sources [6], the various discursive communities are differently exposed to dis/mis-information campaigns.

Following a previous study on a similar data set [6], the discursive communities largely overlap with political parties. As already commented in [6], this behaviour is probably due to the pre-existence of discursive communities that shape and condition the platform environment. Otherwise stated, the debate about the spread of coronavirus and the efficiency of the adopted countermeasures developed in a context in which various discursive communities are already present: if a users is following a group of accounts, those will be the group from which she/he will receive updates and with which she/he will engage in discussions. In the present paper, as well as in previous ones by the same authors, we will avoid the term *echo chamber*, due to its different usage in the literature: while in [42], Twitter echo chambers are defined via the patterns of news consumption (news outlets are given a political orientation and the user is given a membership based on the her/his news consumption), here we follow the approach in [6, 17–19] and avoid any *a priori* labeling; instead, we focus on the interactions among users whence our choice of calling them *discursive communities*.

The presence of politically-oriented discursive communities is reflected in the structure of the discussion, mimicked by the structure of our semantic network. Remarkably, the hashtags referring to d/misinformation arguments represents a minority of the entire semantic network, confirming the observations in [2]. Interestingly, not all users are exposed in the same way to those hashtags: as already observed in [6], right-wing discursive communities use more hashtags related to d/misinformation.

The manuscript is organised as follows: we present the main results in Sect. 2, discuss the various discursive communities in Sect. 2.1 and analyse the semantic network in Sect. 2.2. Then, we briefly present the methodology used in Sect. 3 and conclude by discussing our findings in Sect. 4.

2 Results
2.1 Identification of the discursive communities
Users can interact on Twitter in different ways: for example, one can *retweet* the content of another user, hence endorsing it [43] and raising the content visibility; in order to infer the membership of the various accounts, in the present paper we leverage on this activity, following the procedure adopted in [17, 18].

2.1.1 Discursive communities of verified users
On Twitter there are essentially two kind of accounts: the ones that are verified and whose authenticity is certified by Twitter itself—and belonging to journalists, politicians, VIPs or

being the official accounts of ministries, political parties, newspapers and TV-channels—and the ones that are not verified. About the former ones, we have the largest available information: interestingly enough, verified accounts are more devoted to product original posts than sharing existing ones [6]. Indeed, these accounts act like seeds, proposing new arguments for the public debate.

First, we divided the users of our data set into two groups: the verified and the non-verified accounts. Then, we represented the system as a bipartite network, where verified users are gathered on one layer and the non-verified users are gathered on the other; an edge between vertices of different layers indicates that one has retweeted the other's content at least once during the period of study. To infer the membership of the verified users to a certain discursive community, we projected the bipartite network onto the layer of verified accounts. The procedure consists in counting, for each pair of verified accounts, how many non-verified users have retweeted both of them. The rationale is the following: the largest the number of non-verified users interacting (via tweet or retweet) with the same couple of verified accounts, the greater the possibility that the two are perceived as similar by the audience of unverified ones [17]. Nevertheless, the sole information regarding the number of common neighbors is not enough to state if the two verified accounts are similar: two users may have a great number of common neighbours just because they are both extremely active on Twitter or because their nearest neighbours are among the most active unverified accounts. The statistical significance of the number of common neighbours between two verified accounts can be evaluated by comparing it with its expected value, according to a null-model [32]; once the amount of common nearest neighbours is deemed as statistically significant, we can connect the considered couple of nodes in the projected network.

In the present case, the adopted benchmark is the entropy-based null-model constraining the degree sequences of the bipartite network i.e. the Bipartite Configuration Model (*BiCM* [25]). The details about the whole procedure, i.e. the null-model construction and the validation, can be found in Sect. 3.

The result of the projection is a monopartite network of 3786 edges and 576 different verified users; we used the Louvain algorithm [44] to detect the various communities.[1] The algorithm provided different communities of verified users with an overall modularity equal to 0.61. Discursive communities with a clear political orientation were already detected in other works [17, 18] but the arguments studied there (i.e. the 2018 Italian elections and the political debate about migration policies) were political in nature. Remarkably, as already observed in [6], the political discursive communities shape even the wider debate targeting the Covid-19 pandemic—and including different health, scientific, societal, economic and political facets. In order to gain more insight on the partition and spot the presence of sub-communities, we re-run the Louvain algorithm inside each one of the largest groups: as a result, we individuated five major modules that can be associated with the main Italian political parties (more details on the Italian political scenario and the identity of the verified users mentioned below can be found in Appendix A):

- The **M5S** community contains 85 accounts of supporters and politicians of the *Movimento 5 Stelle* party. This community includes the official account of the

---

[1]Respect to the standard Louvain algorithm, using the definition of the Modularity using the Chung-Lu null-model [45], in the present paper we used the monopartite entropy-based configuration model [23]. Literally, Chung-Lu null-model can be considered as a sparse matrix approximation of the null-model in [23] and, indeed, returns wrong results in the presence of strong hubs [22].

movement and the accounts of personalities like *Beppe Grillo*, *Luigi Di Maio* and *Virginia Raggi*. Interestingly the account of the former premier *Giuseppe Conte* is in this community. There are also some official accounts of ministries (like *Ministero della Giustizia* and *Ministero del Lavoro*[2]) and newspapers and TV-channels like *Il Fatto Quotidiano* and *Report Rai 3*.

- The right-wing (**DX**) community is constituted by the supporters and the politicians of the right-wing parties *Lega Nord* and *Fratelli d'Italia*. This is much smaller than the previous one (only 32 elements) and contains the accounts of *Matteo Salvini*, *Giorgia Meloni*, *Lorenzo Fontana*, *Vittorio Sgarbi* and the *Russian embassy*.
- The Democratic Party (*Partito Democratico*, or **PD**) community contains politicians and supporters of the main center-left party. It contains 37 nodes among which the official accounts of politicians as *Nicola Zingaretti*, *Paolo Gentiloni*, *Enrico Letta* and the party official one (*pdnetwork*).
- The Italia Viva (**IV**) group contains accounts of politicians affiliated to the homonym center-left wing party. Here we can find the official account of the party and the accounts of *Matteo Renzi*, *Ivan Scalfarotto* and *Maria Elena Boschi*. Interestingly, we also signal the presence of *Roberto Burioni*, one of the most popular Italian virologists nowadays, particularly active on the popularization on subjects related to the pandemic. The group contains 24 accounts in total.
- The Forza Italia (**FI**) community contains only 11 accounts and all of them are of politicians affiliated to the *Forza Italia* center-right wing party; for example, it contains accounts like *Silvio Berlusconi*, *Antonio Tajani* and *Renato Brunetta*.

In addition to the political groups described above, we also considered the **MEDIA** community, i.e. the community that contains the official accounts of newspapers (like *Repubblica* and *Agenzia Ansa*), TV-channels (like *La7TV*), radio and other media. This group contains 33 verified accounts.
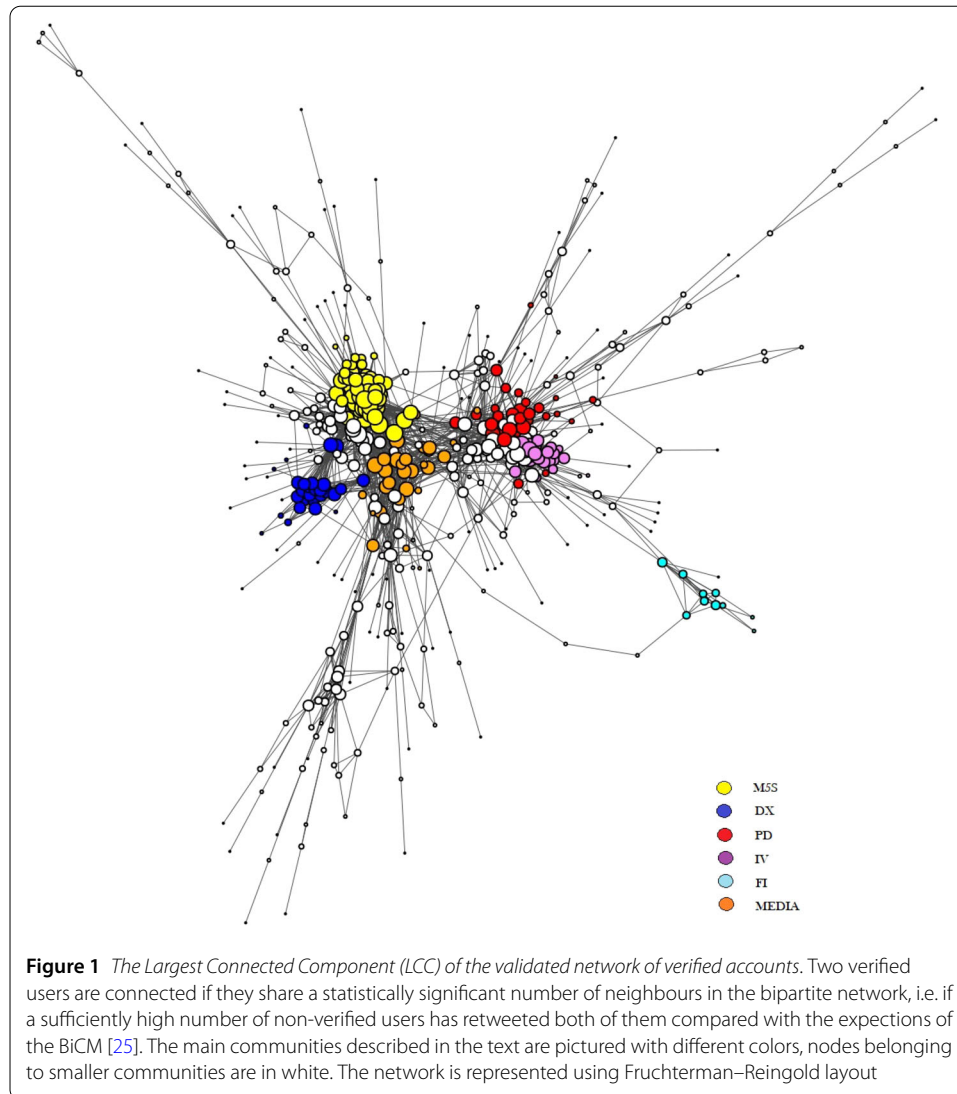
The other discarded discursive communities are mainly very small groups with less than 5 elements. Only three of them result more numerous: the first and bigger one (38 elements) contains accounts of Italian sports journalists and newspapers, football players and clubs (for instance, the official accounts of *Sky Sport*, the football club *AS Roma* or the player *Giorgio Chiellini*); the second one (19 users) consists in accounts related to the *digital world*, like the official account of IBM, TIM or XIAOMI; the last and smaller one (10 elements) includes accounts of Swiss politicians and media. Given their nature, we speculated that these accounts (and people interacting with them) do not actively participate to the political debates created around the topic of the Covid-19 epidemic; therefore, we chose to not include them in our analysis.

The Largest Connected Component (LCC) of the validated network of verified users is shown in Fig. 1. The main communities are depicted with different colors.

2.1.2 Political orientation of non-verified users

Once verified accounts are associated to the various discursive communities, following the approach of. [43], we can infer the membership of unverified ones by considering their interactions in the retweet network. As in [6, 18, 19], we use the membership of verified users as (fixed) seeds for the label propagation proposed by Raghavan et al. [46]. Let us

---

[2]Ministry of Justice and Ministry of Labour.

Figure 1 *The Largest Connected Component (LCC) of the validated network of verified accounts*. Two verified users are connected if they share a statistically significant number of neighbours in the bipartite network, i.e. if a sufficiently high number of non-verified users has retweeted both of them compared with the expectations of the BiCM [25]. The main communities described in the text are pictured with different colors, nodes belonging to smaller communities are in white. The network is represented using Fruchterman–Reingold layout

remind that in case this algorithm cannot find a dominant label for a specific vertex (in case of a tie), it randomly removes some of the edges attached to that vertex and repeats the procedure; due to its intrinsic stochasticity, we run the label propagation 500 times and assigned to each node the most frequent label (actually, the noise in the assignment of the labels is extremely limited): as a result, approximately 89% of the users in the network have been inserted in one of the 6 discursive communities described in the subsection above. They are distributed as follows:

- 117,798 users in MEDIA group;
- 27,989 users in DX group;
- 7230 users in M5S group;
- 1685 users in IV group;
- 1408 users in PD group;
- 430 users in FI group.

As expected, the MEDIA community is the biggest one: it represents users who considerably share news from the accounts of newspapers, radios or newscasts. Looking at the political groups, it is interesting to see that the M5S group contains less elements than the

DX community: by considering only verified accounts, the M5S community includes more than 1.5 times the total number of users of the DX one. The center-left wing (PD and IV) and FI communities are quite small; the vertices with the largest degree are mostly verified accounts, belonging to the MEDIA group (e.g. newspapers or press agencies as *La Repubblica*, *La Stampa*, *Ansa*).

The most retweeted accounts are those of *Giorgia Meloni* (DX community) and *Roberto Burioni* (IV community). Remarkably, the vertex with the largest degree, that the label propagation algorithm assigns to DX community, is a non-verified user whose number of neighbors amounts at 27,000 and whose activity is that of sharing news everyday.[3] It often shares racially-motivated news that the debunking web-site Bufale.net has identified as lacking in sources.
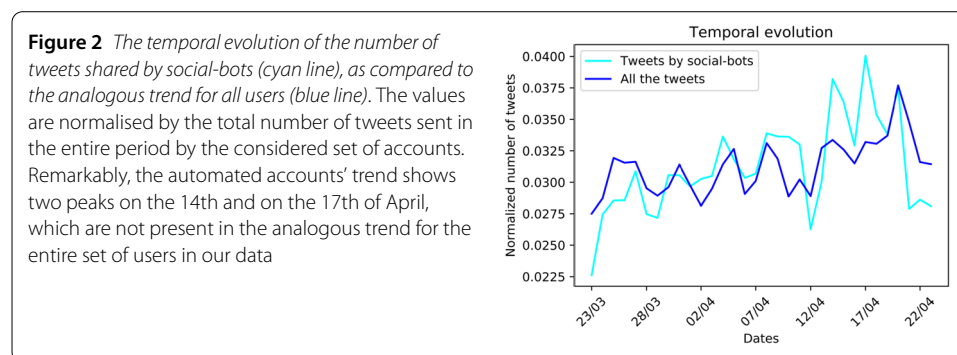
Overall, our results confirm the ones observed in other works, i.e. that communication on Online Social Networks (OSNs) is characterized by a strong polarization, in turn inducing a strongly modular system [6, 16–19, 43, 47–49].

2.1.3 Social-bots

Social bots, or simply bots, are social accounts governed—completely or partly—by pieces of software that automatically create, share and like contents on Twitter and other platforms. In general, the usage of automated accounts is allowed by Twitter platform for promotional purposes by various companies (see Twitter Developer's Automation Rules). Nevertheless, bots often pretend to be human accounts and aim at influencing and diverting the course of discussions by inflating the visibility of some genuine accounts [10, 14].
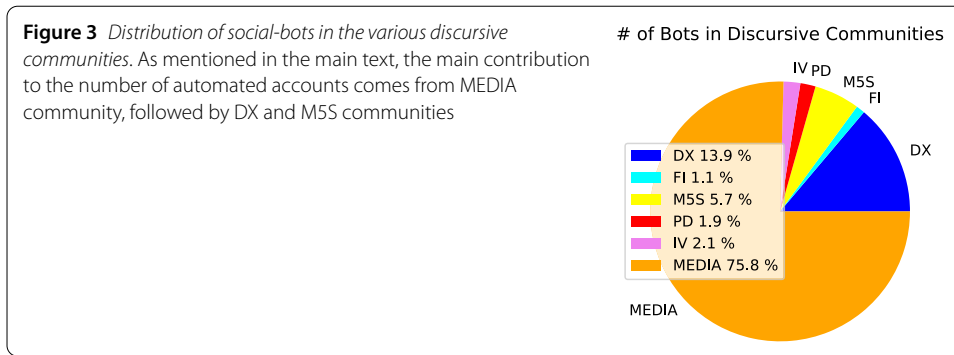
In the present manuscript we used *Botometer* [14], a tool based on supervised machine learning: given a Twitter account, Botometer extracts over 1000 features and produces a classification score called "bot score": according to the algorithm, the higher the score, the greater the likelihood that the account is controlled completely or in part by a software. Botomoter revealed that in our data set social-bots shared 52,054 different tweets (approximately the 3%–4% of the entire data set) with 74,884 hashtags. The most used hashtags by bots are: *#iorestoacasa*, *#italia*, *#news*, *#quarantena* and *#conte*.[4]

In Fig. 2 the temporal evolution of the number of tweets shared by social-bots during the period of study is displayed. Remarkably, this trend is similar to the trend for the entire set of users, except for the presence of two peaks, on the 14th and on the 17th of April.

Figure 2 *The temporal evolution of the number of tweets shared by social-bots (cyan line), as compared to the analogous trend for all users (blue line)*. The values are normalised by the total number of tweets sent in the entire period by the considered set of accounts. Remarkably, the automated accounts' trend shows two peaks on the 14th and on the 17th of April, which are not present in the analogous trend for the entire set of users in our data

---

[3]For privacy reasons, we are allowed to mention the screen name of verified accounts only.

[4]'stay at home', 'italy', 'news', 'quarantine' and Giuseppe 'Conte', the former Prime Minister.

Figure 3 *Distribution of social-bots in the various discursive communities.* As mentioned in the main text, the main contribution to the number of automated accounts comes from MEDIA community, followed by DX and M5S communities

# of Bots in Discursive Communities

- DX 13.9 %
- FI 1.1 %
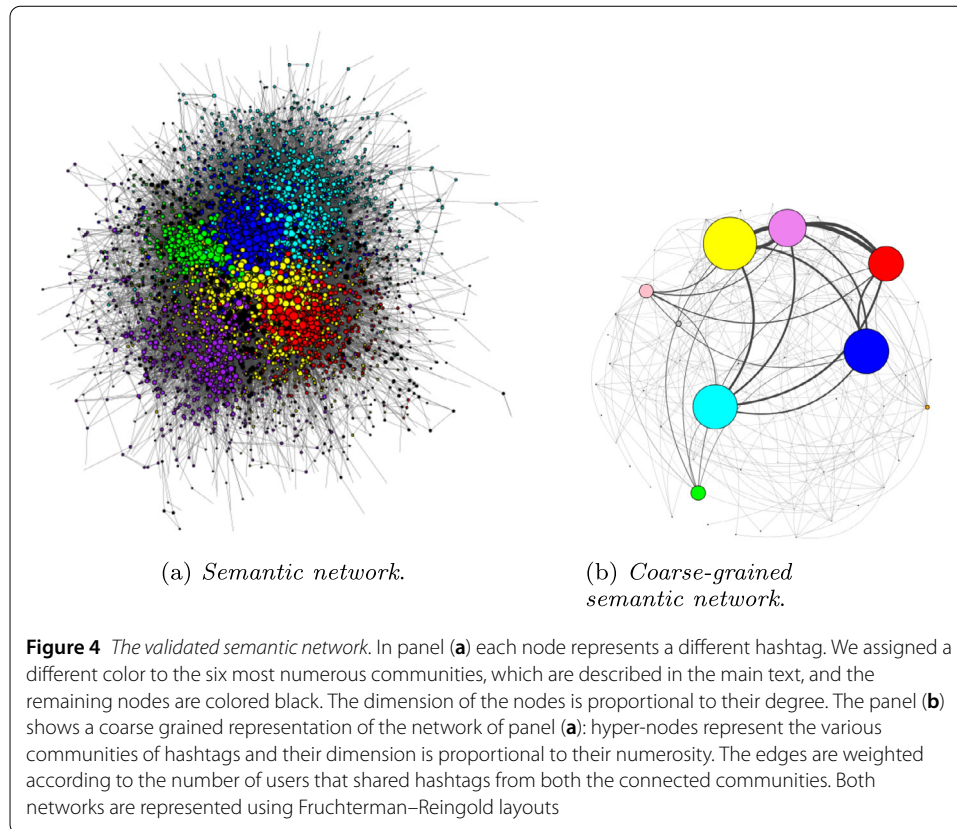- M5S 5.7 %
- PD 1.9 %
- IV 2.1 %
- MEDIA 75.8 %

The intense activity of bots in these two days is linked to the discursive community of MEDIA. Indeed, the 78% and 76% of their retweets in April 14 and 17 respectively, were directed to accounts of this group and only the 16% to users of the DX one (much lower percentages in the case of the other communities). Among the most posted hashtags by bots, there are some references to *Coronabonds*, the prime minister at the time of the data collection Giuseppe Conte and the right-oriented Italian party of *Lega Nord*. Literally, in the mid of April, the political debate in Italy became more intense, in particular about the European Stability Mechanism (ESM). The ESM is an international organization born as a European financial fund for the financial stability of the euro area; in those days the Italian government was considering the possibility of using these funds in order to limit the impact of the pandemic, analysing the possible consequences. In particular, right-, center-right-wing parties and M5S were against its usage, while both PD and IV were in favour of it. In particular, the prime minister Giuseppe Conte proposed to avoid the usage EMS in favour of the European Bonds (or Coronabonds). Lega, still against the usage of ESM, in the European parliament voted against Coronabonds and so it was strongly criticized. The extremely technical nature of the problem could explain the lower interest of the other discursive communities respect to that of the MEDIA group.

Then, we looked at the interactions of social bots with the political groups identified before; in particular, we saw how many bots retweeted contents from verified accounts: while most of the retweeted accounts by bots belong to MEDIA group (e.g. *La Repubblica*, *Agenzia Ansa*, *Sky TG 24*), the most retweeted people are *Roberto Burioni* and *Giorgia Meloni*. In general, bots interact most with the MEDIA group; then, they followed the verified accounts of M5S, DX, IV, PD and FI.

The community with the largest percentage of social-bots is the FI community with 21.9%, followed by IV with 10%, PD with 8.4%, M5S with 6.5%, MEDIA with 5.3% and DX with 4.1%. When absolute numbers are considered, a strong prevalence of bots from MEDIA community appears, followed by those of the DX community, see Fig. 3. More details about the temporal evolution of the activity of automated accounts can be found in the Appendix B.

2.2 The semantic network

Let us now analyse the evolution of the narratives characterizing the online debate during the first peak of the contagion. As in [19, 20], we start from the bipartite network of accounts and hashtags, where a link between the user *u* and the hashtag *h* is present if user *u* has used hashtag *h* at least once. Then, by using the same procedure implemented to

(a) *Semantic network.*   (b) *Coarse-grained semantic network.*

Figure 4 *The validated semantic network.* In panel (**a**) each node represents a different hashtag. We assigned a different color to the six most numerous communities, which are described in the main text, and the remaining nodes are colored black. The dimension of the nodes is proportional to their degree. The panel (**b**) shows a coarse grained representation of the network of panel (**a**): hyper-nodes represent the various communities of hashtags and their dimension is proportional to their numerosity. The edges are weighted according to the number of users that shared hashtags from both the connected communities. Both networks are represented using Fruchterman–Reingold layouts

determine the discursive communities of the validated accounts, we can extract the (validated) semantic network. As already mentioned in the Introduction, the approach that we follow in the present manuscript is slightly different from the one used in [19, 20]: there, the authors analysed the semantic network defined by each discursive community, while here we consider the "global" semantic network and how the various discursive communities interact with it. The resulting network is formed by 5666 different hashtags, linked by 90,560 connections.

Interestingly, even if the main topic is not strictly political, the most connected hashtags, i.e. those with the highest values of the degree, refer to political parties and politicians: *#pd*, *#oms*, *#m5s*, *#lamorgese*, *#regione*, *#lazio*, *#dimaio*, *#governo*, *#zingaretti*, *#mes* and *#conte*.[5]

We, then, run the Louvain algorithm again to detect the various semantic communities; the algorithm provided 61 different communities of hashtags (with modularity $Q \simeq 0.56$). We just focused on the most populated ones (see Fig. 4). The biggest communities refers to some of the most debated themes and subjects during the pandemic, in particular:

- the Red community contains mostly political subjects: here we can find the name of the governing political parties at the moment of the data collection. In this sense, heavy criticisms towards the Prime Minister are present;

---

[5]Respectively, 'Democratic Party', 'World Health Organization' (OMS is the Italian acronym), 'Movimento 5 stelle' (the most represented party in the Italian Parliament), 'Luciana Lamorgese' (Italian interior minister), 'region', 'Lazio' (the Italian region including Rome, administrated by the Italian Democratic Party), 'Luigi Di Maio' (at the time of data collection, Italian foreign minister and leader of the M5S), 'government', 'Nicola Zingaretti' (Democratic Party secretary at the time of data collection and present governor of Lazio), 'European Stability Mechanism' (the above mentioned ESM) and 'Giuseppe Conte', Italian president of the council of ministers at the time of data collection.

- the Violet community includes subjects related to the Catholic Church and to the Pope Francis I;
- the Yellow community is the most crowded one with pieces of news related to either the local (e.g. at regional level) or the global response to the epidemic;
- in Blue, we find updates of the Covid-19 situation (number of deaths, number of contagions, etc.);
- the Cyan community includes hashtags related to trade unions, remote working and to actions adopted by the government at the time of data collections to sustain the employment. Those arguments quickly became hot topics, since the Covid-19 had a heavy impact on the employment and forced firms to take countermeasures such as remote workings [7];
- the Green community includes hashtags related to sports—in particular football, the most followed sport in Italy—that, as many other activities, had to stop.

More details about the hashtags in the various communities of the semantic network can be found in Appendix C.
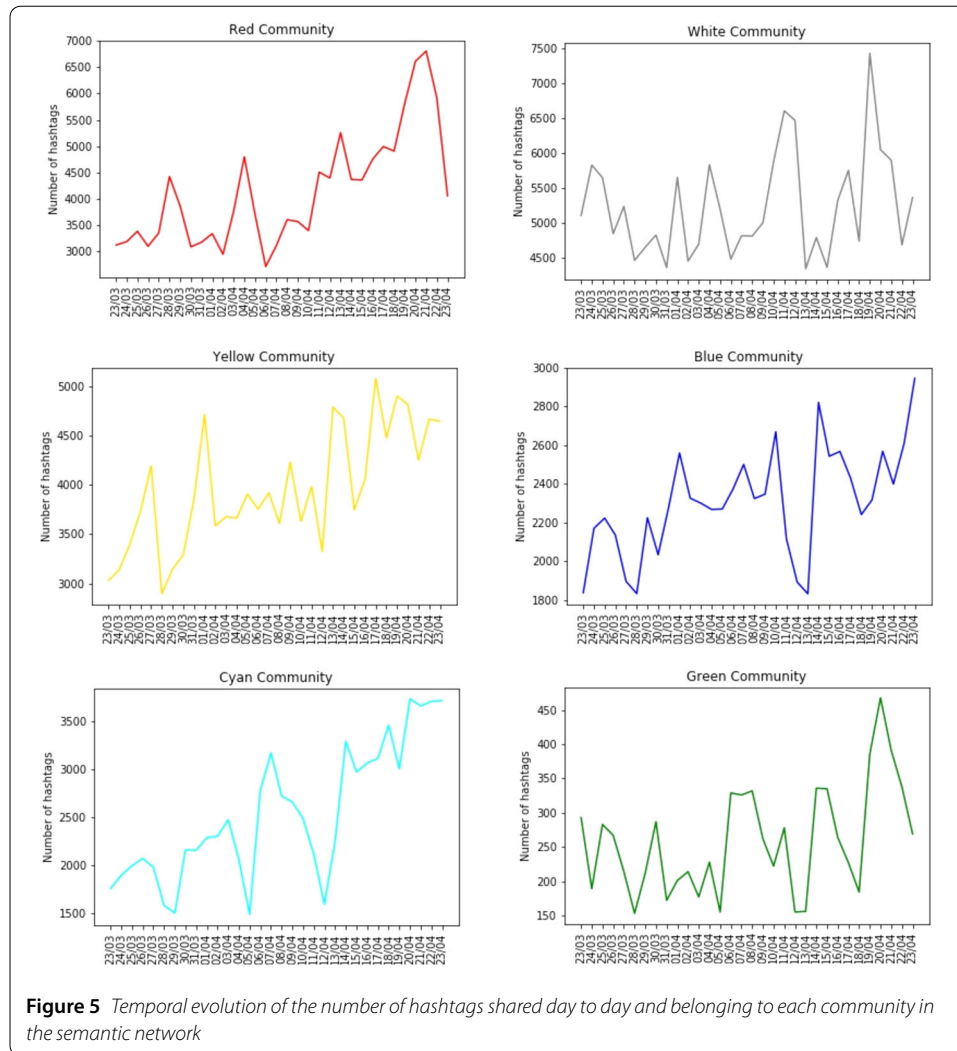
2.2.1 Temporal activities over the semantic network

After the identification of the most important topics within the main communities, we examined the temporal evolution, on a daily scale, of the number of published hashtags belonging to each community (see Fig. 5). Tracking these temporal behaviours is important for understanding which events may have caused an increasing Twitter activity about a specific topic. By looking to the peaks in the temporal evolution, the first thing that catches the eye is that all the trends are upwards, indicating that the Twitter conversations about the Covid-19 became more intense since middle April. We identified some events in specific days, which are strictly related to the main topic of the community in exam:

- **17/04/2020**: in this day there was the vote on the activation of "corona bonds" (joint debt issued to member states of the EU) at the European Parliament. The parties of *Lega Nord* and *Forza Italia* voted against and this caused a lot of comments also on Twitter. There is a peak in this day in the yellow community, which contains also the hashtags #*coronabond*, #*eurobond*, #*lega* and #*salvinisciacallo*.[6]
- **19/04/2020**: in this day there is a peak in the curve of the violet community. In that day the hashtag #*25aprile* (i.e. the Italian liberation day from nazi-fascist occupation) was published many times. In particular, some statements of the senator *Ignazio La Russa*, about the nature of the commemorations in that day, caused debates and controversies on Twitter.[7] Indeed, also the hashtag #*ignaziolarussa* is contained in the violet community.
- **20/04/2020**: the so-called "second phase", during which less severe measures have been implemented and some shops and workplaces reopened, started this day. The cyan community contains the hashtags #*fase2* (phase 2) and consequentially its trend shows a peak on 20/04. In this day a tweet of the US President at the time of data
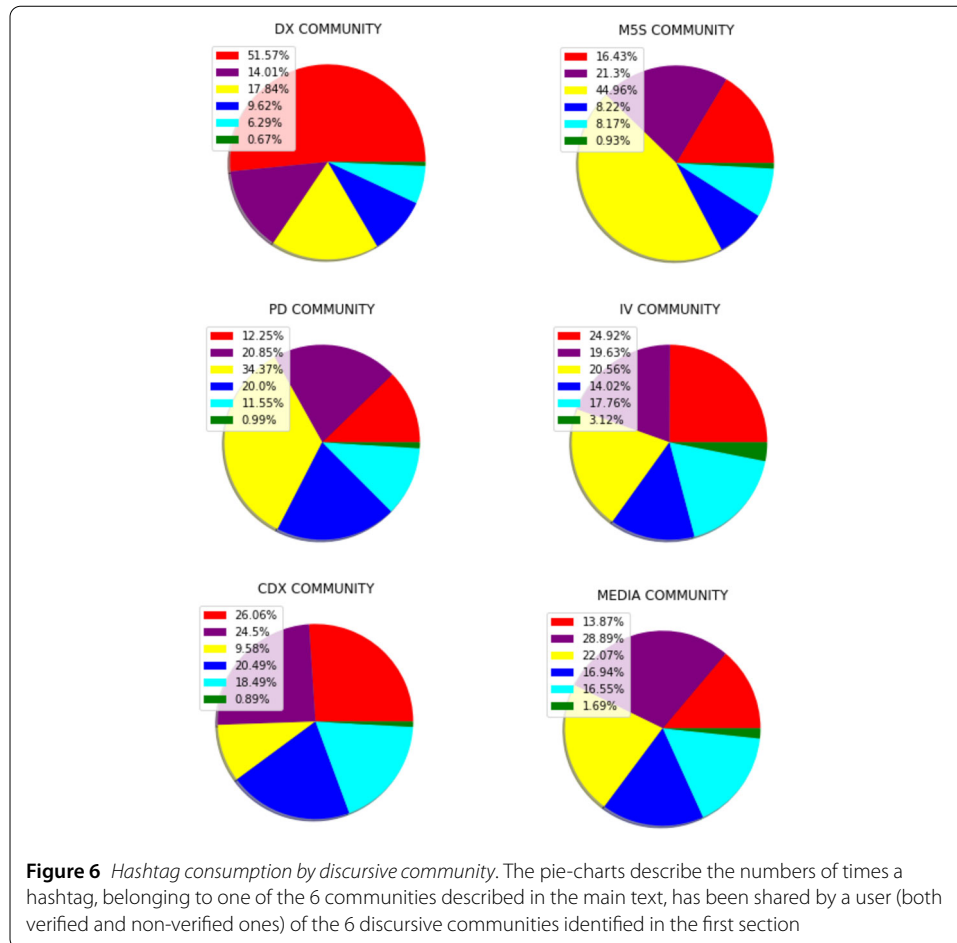
---

[6]Literally, 'Salvini jackal', a clear accusations against the leader of Lega Nord, Matteo Salvini, to take advantage of the critical situation to increase his consensus by causing unnecessary difficulties to the government against the interest of the nation.

[7]The Senator Ignazio La Russa, Fratelli d'Italia, proposed to convert 25th of April from the Liberation (from nazi-fascism) day to the "day of the commemoration of all the victims of all wars, included the one against the Coronavirus". Due to his membership to a right-wing party, this proposal was read as a tentative to minimize the role of nazi-fascist regime in Italy.

Figure 5 *Temporal evolution of the number of hashtags shared day to day and belonging to each community in the semantic network*

collection, i.e. Donald Trump, declared that he would sign an executive order for suspending immigration in the United States to stop the virus and this announcement caused comments and debates on social networks. Again in the yellow community, which contains also the hashtag *#trump*, there is a peak in this day. Also the green community shows a maximum on the 20th of April, due to the statements of Sport Minister Vincenzo Spadafora in which he expressed doubts about the resumption of the Italian football championship Serie A.

- **21/04/2020**: the temporal evolution of the red community, i.e. the "political" one, shows a peak in this day. The most-shared hashtag of this community, in this day, is *#quartarepublica*, an Italian TV-program. The day before, the main host of the program was Silvio Berlusconi, leader of the "Forza Italia" party and former Prime Minister. Other hosts were the Governor of Veneto region Luca Zaia, the Councilor for Welfare of the Lombardy Giulio Gallera and the Mayor of Naples Luigi De Magistris. As observed in similar studies [19, 20], right and center-right wing users are particularly active on mediated events, i.e. public events as TV interview that are heavily covered by users on Twitter.

Figure 6 *Hashtag consumption by discursive community.* The pie-charts describe the numbers of times a hashtag, belonging to one of the 6 communities described in the main text, has been shared by a user (both verified and non-verified ones) of the 6 discursive communities identified in the first section

2.2.2 Semantic activity of the discursive communities

Figure 6 shows the semantic activity of the various discursive communities, including both verified and unverified users. In the DX group (and, similarly, also in the FI group) there is a sharp prevalence of the red community, due to the presence of hashtags against the government. For what concerns the M5S and the PD, the yellow community is the most shared one. Within the MEDIA group, hashtags are homogeneously distributed among violet, yellow, blue and cyan communities.

We repeated the same analysis for social-bots and observed that hashtags by bots are quite evenly distributed across all communities (with the exception of the green one): still, social-bots interact mostly with the verified accounts of MEDIA group and, therefore, they share news and contents of different types. In general the situation changes when considering only verified accounts: more details can be found in Appendix D.
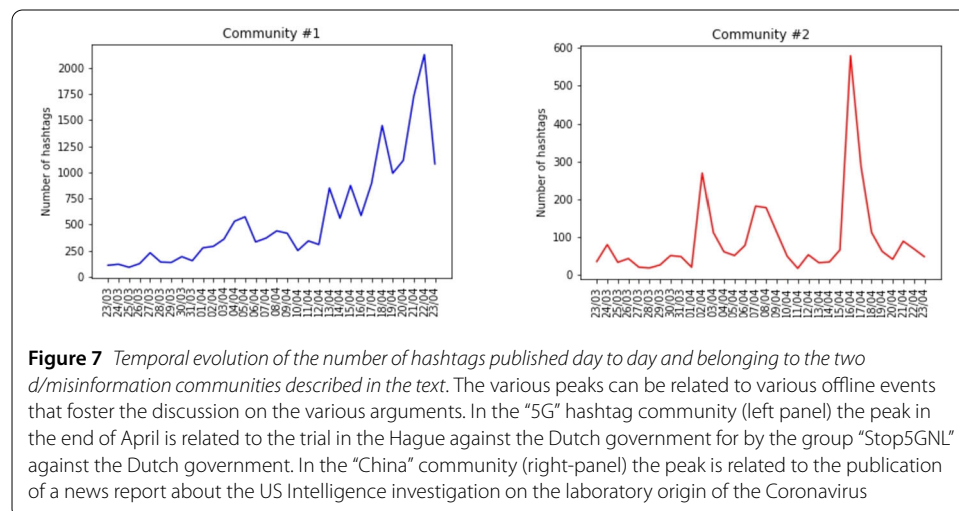
2.2.3 Tracking conspiracy theories and d/misinformation campaigns

In the literature, "disinformation" and "misinformation" have different meanings, both referring to the spread of false information: the former concerns deliberate diffusion, while the latter refers to an unintentional mechanism [50, 51]. At the present level, we cannot distinguish between the two different natures, thus we always use the term d/misinformation. One of the most interesting aspects of our analysis is the possibility of investigating the spread of forms of d/misinformation online. We have identified 2 sub-

communities of hashtags related to d/misinformation campaigns regarding the origin and the diffusion of the coronavirus.
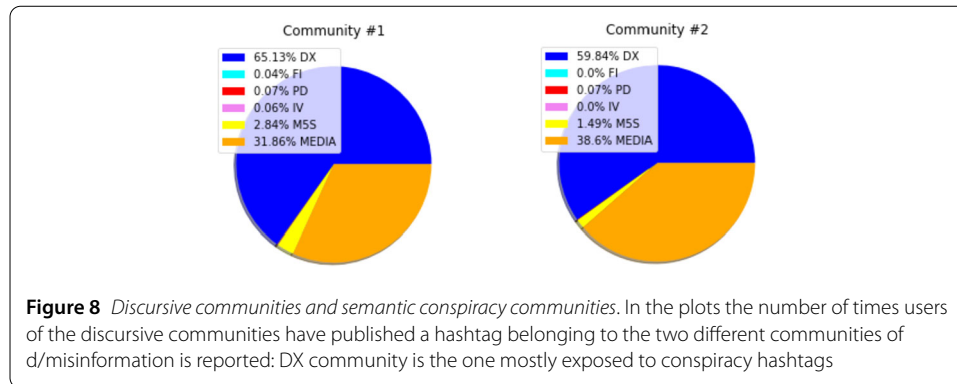
Looking at the hashtags of the first community, connections between Bill Gates, vaccines, 5G, nano-/micro-chips and naturally the Coronavirus, emerge.[8] Indeed, one of the most widespread false claims seems to be the theory for which the pandemic is a plan masterminded by Bill Gates to implant microchips into humans along with a Coronavirus vaccine. Other interesting connections are those between the hashtags *#colao* and *#montagnier*: the first refers to Vittorio Colao, ex CEO of Vodafone and new director of the task force formed by the premier Giuseppe Conte, and the latter refers to the Nobel prize for medicine in 2008, Dr. Montagnier. Conspiracy theorists extracted one phrase from a 2019 video, in which Colao said something about a "medical substance" that could be injected thanks to 5G. Instead, Dr. Montagnier stated in an interview that the spread of Coronavirus was a human error by scientists trying to develop a vaccine, precisely against AIDS (in other sub-communities we found also hashtags like *#HIV* and *#AIDS*). There are also some references to "Immuni" App, which is the application developed by Bending Spoons company, appointed by the Italian government for contact-tracing in order to control the spread of the epidemic; shortly after the release of the app, there were worries about privacy and some users argued that the app was created for spying people.

The second community is about the creation in laboratory of the virus by Chinese scientists.[9] We plotted the temporal evolutions of the daily number of published hashtags, also for these communities in Fig. 7, trying again to identify those events that may have caused an increasing attention on Twitter about these topics. For the community about 5G there is a peak on April 21st and 22nd; on the former there was a trial in the Hague court against the Dutch government for the introduction of 5G brought by the group "Stop 5GNL". Moreover, rumors about Bill Gates started to spread from middle April, when conspiracy theorists used a "TED Talk" from 2015 in which Gates warned that the world is not

Figure 7 *Temporal evolution of the number of hashtags published day to day and belonging to the two d/misinformation communities described in the text.* The various peaks can be related to various offline events that foster the discussion on the various arguments. In the "5G" hashtag community (left panel) the peak in the end of April is related to the trial in the Hague against the Dutch government for by the group "Stop5GNL" against the Dutch government. In the "China" community (right-panel) the peak is related to the publication of a news report about the US Intelligence investigation on the laboratory origin of the Coronavirus

[8]This community includes the following hashtags: *#5G, #5Gkills, #antiinfluenzale, #appimmuni, #bendingspoons, #bigpharma, #bill, #billgates, #billgatesfoundation, #colao, #deepstate, #immuni, #microchip, #montagnier, #nanochip, #no5G, #noimmuni, #spygate, #stop5G, #telecontrollo, #vaccino, #vittoriocolao.*

[9]This community includes the following hashtags: *#china, #chinacentric, #chinaliedpeopledied, #chinamustpay, #chinavirus,#chinesecoronavirus, #coronaviruscure, #harvarduniversity, #trumpderangementsyndrome, #who.*

Figure 8 *Discursive communities and semantic conspiracy communities*. In the plots the number of times users of the discursive communities have published a hashtag belonging to the two different communities of d/misinformation is reported: DX community is the one mostly exposed to conspiracy hashtags

prepared for an epidemic, to confirm their theories. The second community has a peak on the 16th of April when a news report was published about the American intelligence investigating the alleged creation of the Coronavirus in laboratory in Wuhan.

Confirming the results of Ref. [2], those conspiracy communities represent a minority in the semantic network: the hashtags included in the conspiracy communities are respectively 62 and 15, over a total of 5666 different hashtags. Even in terms of their popularity, their impact is limited: the first conspiracy community was shared 10,452 times and the second one 1287, against a total of nearly 602,299 messages containing at least one hashtag.

Remarkably, not all discursive communities share conspiracy hashtags in the same way. In Fig. 8 there are the fractions of users of the 6 different discursive communities listed in the previous paragraph, which shared the hashtags of the three communities of d/misinformation. The DX community is the one most affected by d/misinformation, followed by MEDIA group which however contains much more users than the DX one. Moreover, the discursive communities not included in our analysis do not participate at all to the sharing of d/misinformation contents. Even in this sense, our analysis confirms the findings of Ref. [6]: there the Non Reliable sources, as tagged by NewsGuard, were almost exclusively shared by DX community (in Ref. [6] MEDIA community was not analysed).

3 Methods

3.1 Data

The data set analysed is similar to the one used in [6]. Data were collected using Twitter Search API across a period of 32 days, between the 23th of March 2020 to the 23th of April 2020, that was one of the most crucial period of the quarantine. In particular we looked at those tweets containing at least one of the following keywords or hashtags: *coronavirus, coronaviruses, ncov, ncov2020, ncov2019, covid2019, covid-19, SARS-CoV2, #coronavirus, #WuhanCoronavirus, #coronaviruschina, #CoronavirusOutbreak, #coronaviruswuhan, #ChinaCoronaVirus, #nCoV, #coronaviruses, #ChinaWuHan, #nCoV2020, #nCoV2019, #covid2019, #covid-19, #SARS_CoV_2, #SARSCoV2, #COVID19.*

The total number of tweets selected in this way was approximately 1.5 millions, posted by 126,614 different users. Then, we also took all the hashtags in each tweet for identifying the central topic. As a consequence, only tweets containing, at least, one hashtag have been retained, resulting in 602,299 messages. In presence of retweets, we looked also for the user IDs and for the hashtags of the original tweets.

At this point, we had to deal with the problem of orthographic errors or, in general, with all the possible variations of a word that have to be considered as an unique one. For this reason we had to clean our data set, identifying, in some way, those hashtags that were "similar enough". We used the Levenshtein distance, that is a sequence-based similarity: it quantifies the cost of transforming a string x into a string y when the two strings are viewed as sequences of characters. In particular we counted the number of characters that had to be changed and we divided it for the total number of characters of the words in exam. We have set a threshold at 0.82 to consider any two strings as the same and kept the one with larger occurrence in the data set. We chose this threshold after several sample checks; this value seemed to be the most effective to identifying the different versions of the same word, reducing the number of different hashtags by about 30%. Afterwards, we put the information about the ID of the users and their hashtags into the two layers of a bipartite network; a link between a hashtag and an user exists when that user has published a tweet containing that hashtag at least once.

3.2 Entropy-based null-models as benchmarks

Once the information about the hashtag usage of the system is represented via a bipartite network, we need to extract signals that cannot be explained only invoking a random usage of hashtag by users. In this sense, we need a proper benchmark: entropy-based null-models are, by construction, unbiased; so, they represent a natural choice [22]. In the present case, we want to extract the various narratives, as identified by group of hashtags, that cannot be explained only 1) by the attitude of users to use different hashtags and 2) by the virality of the latter ones. Since this information is encoded into the degree sequence of both layers, we need a benchmark discounting the information carried by the degree sequence. The entropy-based null-model discounting it is the Bipartite Configuration Model [25]). In the present section we will briefly revise the steps to define this null model.

Let us consider a bipartite network in which the two layers $\top$ and $\perp$ have dimension $N_\top$ and $N_\perp$; in the following, Latin indices will be used to identify nodes on the $\top$ layer while Greek ones will be used for the $\perp$ layer. Then, the bipartite network can be represented by its biadjacency matrix, i.e. a $N_\top \times N_\perp$ matrix $\mathbf{M}$ whose generic entry $m_{i\alpha}$ is 1 if the node $i \in \top$ is connected to the node $\alpha \in \perp$.

First, let us define a *statistical ensemble* of graphs, i.e. the set of all the possible bipartite graphs having the same number of nodes but with different topology, from the fully connected to the empty ones. Then, we can define the Shannon entropy over the ensemble, by assigning a different probability to each element of it:

$$S = - \sum_{G_{\mathrm{Bi}} \in \mathcal{G}_{\mathrm{Bi}}} P(G_{\mathrm{Bi}}) \ln P(G_{\mathrm{Bi}});$$

here, $P(G_{\mathrm{Bi}})$ is the probability of the bipartite graph G_{Bi}. Let us now maximise the entropy, while constraining the network degrees: in particular, we want that the ensemble average of degrees to match their observed value, in order to have a null-model tailored on the real system. In term of the biadjacency matrix, the degree sequences of the $\top$ and $\perp$ layers respectively read $k_i = \sum_\alpha m_{i\alpha}$ and $k_\alpha = \sum_i m_{i\alpha}$. Using the method of the Lagrangian multipliers, constrained maximisation can be expressed as the maximisation of S', defined

as

$$
\begin{aligned}
S' = S \\
+ \sum_i \eta_i \left[k_i^* - \sum_{G_{\mathrm{Bi}} \in \mathcal{G}_{\mathrm{Bi}}} P(G_{\mathrm{Bi}}) k_i(G_{\mathrm{Bi}}) \right] + \sum_\alpha \theta_\alpha \left[h_\alpha^* - \sum_{G_{\mathrm{Bi}} \in \mathcal{G}_{\mathrm{Bi}}} P(G_{\mathrm{Bi}}) h_\alpha(G_{\mathrm{Bi}}) \right] \\
+ \zeta \left[\sum_{G_{\mathrm{Bi}} \in \mathcal{G}_{\mathrm{Bi}}} P(G_{\mathrm{Bi}}) - 1 \right],
\end{aligned}
$$

where S is the Shannon entropy defined above, η_i, θ_α are the Lagrangian multipliers relative to the degree sequences and ζ is the one relative to the probability normalization; quantities marked with an asterisk $*$ indicate quantities measured on the real network.

Maximising S' leads to a probability per graph $G_{\mathrm{Bi}} \in \mathcal{G}_{\mathrm{Bi}}$ that can be factorised in terms of the probabilities per link $p_{i\alpha}$ [52], i.e.

$$
P(G_{\mathrm{Bi}}) = \prod_{i,\alpha} p_{i\alpha}^{m_{i\alpha}(G_{\mathrm{Bi}})} (1 - p_{i\alpha})^{1 - m_{i\alpha}(G_{\mathrm{Bi}})}, \tag{1}
$$

where $p_{i\alpha} = \frac{e^{-\eta_i - \theta_\alpha}}{1 + e^{-\eta_i - \theta_\alpha}}$. Nevertheless, at this level the above equation is just formal, since we do not know the numerical value of the Lagrangian multipliers. To this aim, we can then maximise the likelihood of the of the real network [23, 53]; it can be shown that the likelihood maximisation is equivalent at imposing

$$
\langle k_i \rangle = k_i^*, \quad \forall i \in \top; \qquad \langle h_\alpha \rangle = h_\alpha^*, \quad \forall \alpha \in \bot.
$$

Let us add a final remark. The BiCM is not the only null model properly discounting the information contained into the degree sequence. The *Curveball* approach of Strona et al. [54] is equally ergodic [55], but implements a *microcanonical* approach, i.e. each member of the corresponding ensemble satisfies the degree sequence *exactly* (the BiCM, instead, constrains the degree sequence *on average* over the ensemble). It has been shown that the two approaches, even if employing the same information, are not equivalent [56–58]. In particular, the z-score associated with the observation of non-linear quantities can assume opposite signs under the BiCM or the Curveball, so leading to opposite conclusions: in fact, the presence of fluctuations in the BiCM can screen the real behaviour [58]. In this sense, the Curveball algorithm must be preferred when the data are error-free, since it prevents fluctuation from hiding the genuine behaviour of the system. Instead, when data are subject to errors or samplings (as in the case of Twitter data), fluctuations should be admitted and the BiCM preferred.

3.3 Validated projection of bipartite networks
Once we have a well-grounded benchmark, we want to infer similarities among nodes on the same layer. To quantify the similarity between any two nodes, we employ the number of common neighbours (e.g. the number of users that have shared the same pair of hashtags). These patterns, named *V-motifs*, are members of a wider class of structures [32, 59] intended to capture the topological features of bipartite networks; although other motifs-like subgraphs could have been employed to proxy the similarity between any two vertices activity [32], we have opted for the simplest, yet informative, choice. Moreover, as we will

see below, this choice allows us to rephrase our projection algorithm as an exact statistical test.

Let us assume, without loss of generality, that we want to project the information contained in the bipartite network onto the $\top$ layer and call V_{ij} the number of common neighbors between nodes $i, j \in \top$.[10] In terms of the biadjacency matrix, V_{ij} can be expressed as

$$V_{ij} = \sum_{\alpha} V_{ij}^{\alpha} = \sum_{\alpha} m_{i\alpha} m_{j\alpha},$$

where we have defined $V_{ij}^{\alpha} = m_{i\alpha} m_{j\alpha} = 1$, if both i and j are connected to node $\alpha \in \bot$. Let us now compare the observed numbers of co-occurrences between each possible pair of nodes in $\top$ with the prediction of the BiCM. Since link probabilities are independent, the presence of each V-motif V_{ij}^{α} can be regarded as the outcome of a Bernoulli trial:

$$f_{\mathrm{Ber}}\left(V_{ij}^{\alpha} = 1\right) = p_{i\alpha} p_{j\alpha},$$

$$f_{\mathrm{Ber}}\left(V_{ij}^{\alpha} = 0\right) = 1 - p_{i\alpha} p_{j\alpha}.$$

In general, the probability of observing $V_{ij} = n$ can be expressed as a sum of contributions, running on the n-tuples of considered nodes (in this case, the ones belonging to the layer of users). Indicating with A_n all possible nodes n-tuples among the layer of $\bot$, this probability amounts at

$$f_{\mathrm{PoissBin}}(V_{ij} = n) = \sum_{A_n} \left[\prod_{\alpha \in A_n} p_{i\alpha} p_{j\alpha} \prod_{\alpha' \notin A_n} (1 - p_{i\alpha'} p_{j\alpha'}) \right], \qquad (2)$$

where the second product runs over the complement set of A_n, that is the set of users not belonging to the tuple chosen. Equation (2) represent the generalization of the usual Binomial distribution when the single Bernoulli trials have different probabilities, also known as Poisson Binomial distribution [60].
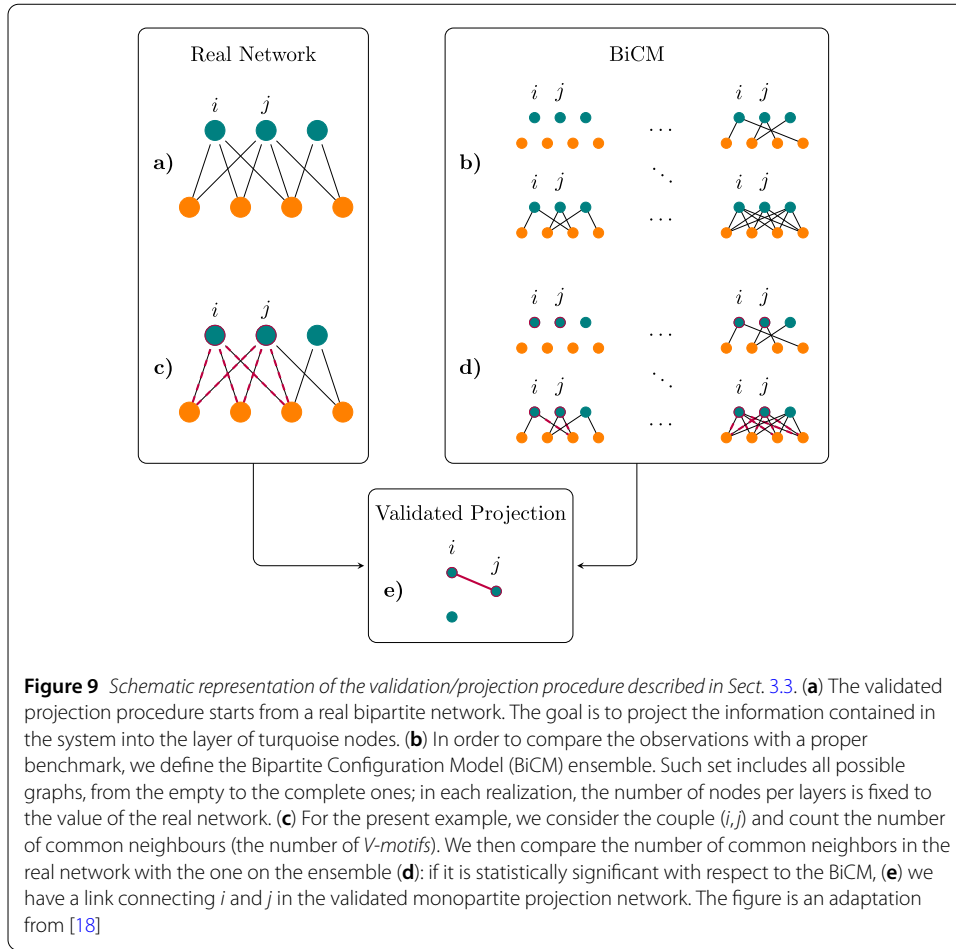
We can, then, verify the statistical significance of the observed co-occurrences by calculating their p-value according to the distribution in (2), i.e. the probability of observing a number of co-occurrences greater than, or equal to, the observed one:

$$\text{p-value}\left(V_{ij}^*\right) = \sum_{V_{ij} \geq V_{ij}^*} f_{\mathrm{PoissBin}}\left(V_{ij}^*\right); \qquad (3)$$

by repeating this calculation for every pair of nodes, we obtain $\binom{N_\top}{2}$ p-values. In words, our validation procedure allows us to filter out signals that are spurious in nature either because a mere consequence of the users activity or because a mere consequence of the hashtags popularity (both effects being proxied by the nodes degrees defining our benchmark model).

In order to assess the statistical significance of the hypotheses belonging to this group, a multiple hypothesis testing correction must be adopted; in the present paper, we use the *False Discovery Rate* (FDR, [61]), since it allows the group percentage of false positives to

---

[10]Following Ref. [32], we use the letter V to indicate common neighbours, since this pattern appear in the bipartite network as a V between the layer.

Figure 9 *Schematic representation of the validation/projection procedure described in Sect.* 3.3. (**a**) The validated projection procedure starts from a real bipartite network. The goal is to project the information contained in the system into the layer of turquoise nodes. (**b**) In order to compare the observations with a proper benchmark, we define the Bipartite Configuration Model (BiCM) ensemble. Such set includes all possible graphs, from the empty to the complete ones; in each realization, the number of nodes per layers is fixed to the value of the real network. (**c**) For the present example, we consider the couple (*i,j*) and count the number of common neighbours (the number of *V-motifs*). We then compare the number of common neighbors in the real network with the one on the ensemble (**d**): if it is statistically significant with respect to the BiCM, (**e**) we have a link connecting *i* and *j* in the validated monopartite projection network. The figure is an adaptation from [18]

be tuned. A pictorial representation of the validation/projection procedure is provided in Fig. 9.

3.4 Modularity and community detection

In the present analysis, we inferred the discursive communities from the communities in the validated network of verified users. In particular, we used the modularity based Louvain algorithm [44].

The modularity [62] compares the number of edges within the actual communities with the number of edges one would expect to have in the same communities but regardless of communities structure; the latter quantity depends critically on the chosen null-model. Modularity can be written as

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(C_i, C_j), \tag{4}$$

where m is the total number of links of the network, A_{ij} are the entries of the adjacency matrix, P_{ij} is the probability to have a link between nodes i and j according to the chosen null-model and the term $\delta(C_i, C_j)$ selects all the pairs of nodes contained in the same community (equal to 1 if $C_i = C_j$ or 0 otherwise). In the original definition in [63] the null model chosen is the Chung-Lu one [45], which conserve the degree sequence, but it is

known to be inconsistent for dense networks that present strong hubs. In the present paper we use instead the entropy-based Undirected Binary Configuration Model (UBCM) defined in [23, 53]: it can be shown that in the case of sparse network, the UBCM can be approximated by the Chung-Lu null-model. In the present case, we implemented the Louvain algorithm using the 'full' null-model, implemented within the Python module `NEMtropy`, described in [64].

As a last observation, we would like to stress that our dataset is the one considered in [6] and that the methodology employed to analyse it is the same. Nevertheless, here we have opted for the 'full' BiCM, implemented within the Python module `NEMtropy`, described in [64]; in [6], instead, the sparse network approximation was used. This may let small disagreements in the composition of the communities found in the present manuscript and of those found in [6] to emerge. In particular, the right-wing community in [6] had more elements and also included the supporters of the *Forza Italia* party—now separated from those of *Lega Nord* and *Fratelli d'Italia*.

4 Discussion

In the present manuscript, we analysed the narratives emerged on Twitter during the Covid-19 epidemic in Italy. Our data set covered a period of one month (from the 23rd of March 2020 to the 23rd of April 2020) and included approximately 1.5 millions of tweets. As observed in previous studies [6], even in the (non strictly) political discussion concerning the Covid-19, the discursive communities reflect political groups. This behaviour is probably due to the pre-existence of discursive communities: if the accounts followed by a specific users have a certain political orientations, this initial bias will influence even other—even non necessarily political—discussions, just because Twitter platform will display on the user home the updates coming from her/his friends.

In the present paper, we have refined the partition in communities obtained in [6] by using the full BiCM, implemented into the Python module `NEMtropy` and described in [64], instead of its approximation; not surprisingly, the recovered partition is reflected into the organization of the online debate.

We represent it as a bipartite network of users and hashtags, where a link connects a user to an hashtag if the former one has used the latter one at least once during the observation period. Then, using a bipartite entropy-based null-model [25] as a benchmark, we have projected the original bipartite network onto the layer of hashtags [26]. Our approach is similar to the one used in [19, 20] but with a crucial difference: there, the various communication strategies developed by the different groups during political events were studied by obtaining a semantic network per discursive community; here, we have limited ourselves to consider just one semantic network for the entire debate.

As observed in [19, 20], the various discursive communities have a different behaviour: in Fig. 6 we showed that, while the discursive communities closer to the government at the time of the data collection (i.e. M5S, PD and IV) focus more on the news related to the pandemic, the ones closer to the opposition (i.e. DX and FI) focus more on political themes, with the aim of highlighting the inefficiency of the measures adopted by the government to contrast the pandemic. Other behaviours of the various discursive communities observed in [19, 20] are confirmed by the present study. For instance, *mediated* events stimulate the discussions on Twitter more in DX and FI community: by looking at the temporal evolution of the communities of the semantic network, we were able to identify these events, as TV-appearances of politicians.

Confirming other previous studies [2], the communities of hashtags related to d/misinformation represent a limited number of hashtags. Nevertheless, it is striking that they are mostly "visited" by a single discursive community, i.e. the right-wing one, as shown in Fig. 8. Indeed, such an impressive percentage of accounts using those hashtags in their messages is not justified by the number of accounts in this discursive community. Such an exposure to d/misinformation campaigns of this discursive community has already been observed in [6]. Remarkably, in this discursive community the incidence of automated accounts is extremely limited, according to Botometer [65]. In a sense, social bots seem to be more focused on "debating" subjects of wide interest: they formed the 3%–4% of the tweets of the data set and the hashtags they share seem to be distributed quite evenly across the different semantic communities. In fact, they mostly retweeted the contents of MEDIA group; their temporal activity do not suggest particular events which could stimulate their activity on Twitter.

Even if the literature is limited to a few studies, it has been shown that often non trivial correlations between offline and online behaviours occur [66–71]. In this sense, the identification of the various communication behaviours of the Twitter users represents a data-driven tool to measure—and, therefore, elaborate different strategies to limit—the exposure of the users themselves to d/misinformation campaigns. In fact, the presence of groups that are particularly susceptible to low-quality contents limits the efficiency of debunking activities [49]: therefore different strategies should be elaborated to effectively fight d/misinformation to avoid the diffusion of dangerous conducts during the pandemic.

Appendix A: Description of the political communities

The reader not used to the Italian political scenario may need a more detailed description of the cited political parties and of the various characters.

The *Movement 5 Stars* (M5S) is a political party founded on the web in 2009 by the comedian Beppe Grillo as an anti-establishment movement. After getting important results at the local elections (like the election of Virginia Raggi as major of the city of Rome and the election of Chiara Appendino as major of Torino in 2016) the M5S resulted as the most voted party, under the guidance of Luigi Di Maio, at the national elections of 2018. Beside the great result (it was the second time that the M5S was present at a national election), the Movement did not have enough representatives in the parliament to form a government. The political group decided to form a post-election alliance with Lega, the most voted political party of the most voted alliance, i.e the right and center-right one. The government was guided by the prime minister Giuseppe Conte and lasted 1 year and 3 months; in August 2019, the leader of Lega Matteo Salvini to negate the confidence vote to Conte and opened a political crisis. After few weeks of debate, the M5S allied with the Italian Democratic Party (*Partito Democratico*, or PD) to form a novel government, after the openness of the former PD secretary Matteo Renzi. The new government was again guided by Giuseppe Conte; Luigi Di Maio, the leader of the Movement, took the role of Minister of Labour. Another important ministry that was ruled by the Movement was the Ministry of Justice, guided by Alfonso Bonafede. From its foundation, the Movement 5 Stars has been supported by the newspaper *Il Fatto Quotidiano*, guided by Marco Travaglio.

The main *right-wing parties* in the Italian parliament are Lega and Fratelli d'Italia, guided respectively by Matteo Salvini and Giorgia Meloni. Together with Forza Italia, a center-

right-wing party guided by former prime minister Silvio Berlusconi), they presented as a right-wing alliance at the national election of 2018. The alliance was the most voted one, but the number of representatives in the parliament was not enough to express a government. Lega, that resulted as the most voted party in the most voted alliances (even if not the most voted party in general), decided to break the previous alliance to make a new one with M5S. The government was guided by Giuseppe Conte; quite obviously, the choice of Matteo Salvini was heavily criticized by the former allies. Lega abandoned the new alliance with M5S in summer 2019 and went to the opposition.

The *Italian Democratic Party* (PD), guided, at the time of the data collection, by Nicola Zingaretti, was the second most voted party at the election at the national elections of 2018. After the alliance with the M5S in summer 2019 to form a new government, PD experienced the split of a group of representatives, guided by the former PD secretary Matteo Renzi, forming the new political party of Italia Viva. Paolo Gentiloni, former prime minister from the end of 2016 to June 2018 and PD parliament representative, was then elected as European Commissioner for Economy. In the first months of 2021, Nicola Zingaretti resigned as PD secretary and was replaced by Enrico Letta.

Italia Viva is the political party founded by the former prime minister and former PD secretary Matteo Renzi, soon after the formation of the second government guided by Giuseppe Conte: Italia Viva continued supporting (with two ministers) the government, but marked its difference from PD. Important personalities in Italia Viva are Maria Elena Boschi and Ivan Scalafarotto. As already demonstrated in previous experiences in PD, Matteo Renzi is a capable user of the new technologies as a form of communication.

Forza Italia (FI) is the political party founded by the former three-times prime minister Silvio Berlusconi. The performances at the last national elections were particularly unsatisfying for FI, that lost its supremacy in the center-right and right-wing alliance in favour of Lega. Important representative of the party are Renato Brunetta and Antonio Tajani, the national coordinator and former President of the European Parliament.

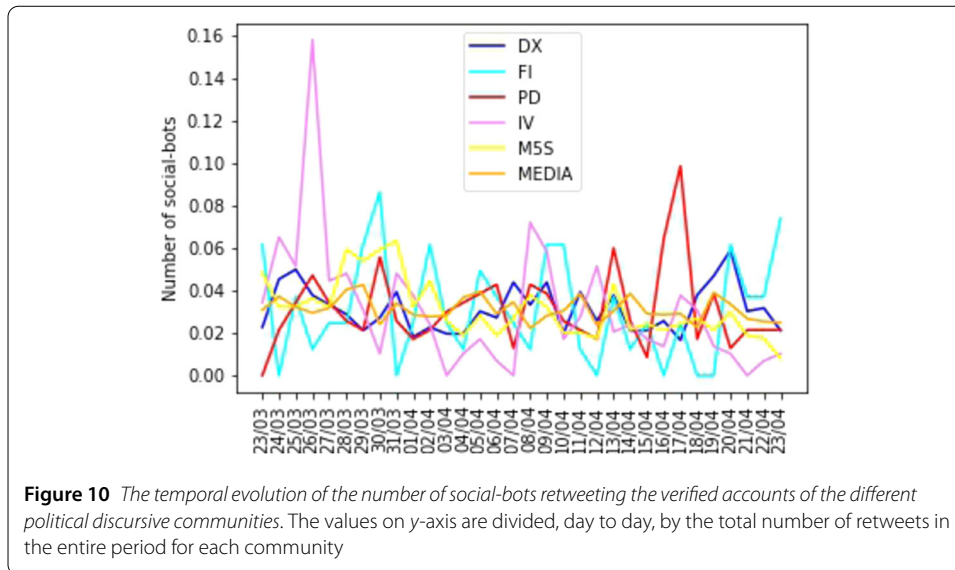## Appendix B: Temporal activity of automated accounts per community

In Fig. 10 the temporal evolution of the number of times bots retweeted a post by one of the six communities of verified accounts is display. The various trends are normalized by dividing for the total number of retweets for each group in the entire period. It is interesting to note the peak on the 26th of March for IV group; in this day bots retweeted many times posts by the virologist Roberto Burioni. Apart from others smaller peaks, the number of bots retweeting verified accounts oscillates around the same values, for all the groups, not suggesting particular events which could have stimulated their activity on Twitter.

Appendix C: Description of the communities in the semantic network

In the present appendix we present a more detailed description of the main communities in the semantic network of Fig. 4.

Red The Red community contains 761 different hashtags. The hashtags with the highest degree are names of politicians or parties, like *#pd*, *#dimaio*, *#lamorgese* and *#zingaretti*.[11] Using again the Louvain algorithm within this community,

---

[11]Democratic Party, Italian foreign Minister Luigi Di Maio, Italian interior Minister Luciana Lamorgese and Democratic party secretary Nicola Zingaretti.

Figure 10 *The temporal evolution of the number of social-bots retweeting the verified accounts of the different political discursive communities.* The values on *y*-axis are divided, day to day, by the total number of retweets in the entire period for each community

we were able to identify specific topics; for instance, in one sub-community we found many hashtags about *immigration* (like *#migranti*, *#immigrati*, *#clandestini*, *#alankurdi*, *#ong*[12]), which has been one of the most discussed political topics in Italy. Especially during the epidemic some extreme right accounts connected the migration from Northern Africa to the spread of the contagions, without reasonable evidences. There is also a sub-community which contains hashtags of protest against the government, like *#governocriminale*, *#governodelcontagio*, *#contedimettiti*, *#contebugiardo* and *#vogliamovotare*.[13] Another popular topic is the European Stability Mechanism (the Italian acronym is MES), which is a European Union agency that provides financial assistance to European countries, in the form of loans or as new capital to banks in difficulty.[14] Indeed, among the most connected hashtags in this community there are also *#esm*, *#mes*, *#sure*,[15] *#nomes* and *#nosure*.

Violet This communities contains 815 elements and the hashtags with the highest degree are all the dates of the period of study (*#23marzo*, *#24marzo* and all the others). We identified two main topics in it: there is a group of hashtags referring to the Pope and the Catholic Church in general (*#papafrancesco*, *#papa*, *#messa*, *#vaticano*, *#chiesa*[16] and others) and another group of hashtags about TV-programs and films (*#tvtime*, *#amici19*, *#piratideicaraibi*, *#harrypotter*[17] and others).

---

[12]'Migrants', 'immigrants', 'Illegal immigrants', the ship 'Alan Kurdi' for the rescue of migrants, 'Non-Governmental Organization'.

[13]'Criminal government', 'Government of the contagion', 'Conte resign', 'Conte liar' (the premier Giuseppe Conte), 'We want to vote'.

[14]https://www.bbc.com/news/business-19870747

[15]The European instrument for temporary "Support to mitigate Unemployment Risks in an Emergency" is intended "to mobilise significant financial means to fight the negative economic and social consequences of the coronavirus outbreak on their territory" (https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/financial-assistance-eu/funding-mechanisms-and-facilities/sure_en).

[16]'Pope Francis', 'pope', 'Mass', 'Vatican', 'church'.

[17]'TV time', the Italian TV-program 'amici', the movies 'Pirates of the Caribbean' and 'Harry Potter'.

Yellow   This community has 1159 elements. It is a very heterogeneous community and it was not easy to understand the common topic. Therefore we computed the betweenness centrality for all the nodes of this community and the hashtag *#regione* ('region') results the most central. Indeed, there are several references to Italian regions like *#basilicata*, *#calabrialeghista*, *#regionesardegna*, *#regionelombardia*[18] and others. In particular most of the hashtags refers to the management of the Coronavirus emergency in Lombardy, the most affected region by the virus in the first wave of the contagion: *#lombardia*, *#commissariatelalombardia* (referring to a petition started on Twitter asking the government to put Lombardy under external administration), *#gallera*, *#fontana*, *#fontanadimettiti*, *#rsa* (in Lombardy the so-called 'Residenze Sanitarie Assistenziali' were violently attacked by the virus), *#milano*[19] and others. Another set of vertices seems to refer to foreign countries and cities: *#francia*, *#cina*, *#usa*, *#giappone*, *#londra*, *#regnounito*, *#newyork*, *#australia*, *#russia*, *#trump*, *#sanchez*[20] and others.

Blue   This community contains 975 hashtags. It collects those hashtags referring to data and updates about Covid-19. Hashtags with the highest degrees are: *#casi*, *#controlli*, *#decessi*, *#ricoveri*, *#calano*, *#positivi*[21] and many others. There are also references to the rules to follow in those days like *#spostamenti*, *#chiusi*, *#comuni*, *#misure*, *#negozi*, *#ordinanza*, *#sindaco*, *#sanzioni*[22] and others. Another common topic seems to be economics and therefore all the measures taken and the decrees issued during the pandemic; some of the most connected hashtags are *#economia*, *#mercati*, *#aziende*, *#credito*, *#finanziamenti*, *#imprese*, *#liquidità*, *#decretoliquidità*[23] and others.

Cyan   In this community with 966 elements, the main topic seems to be about the employment crisis during the pandemic. Among the hashtags with high degrees there are *#lavoro*, *#lavoratori*, *#sindacati*, *#ciglcisluil*, *#pensioni*, *#imprese*.[24] Another topic linked to the previous one regards digital platforms, new technologies and remote working: *#digitale*, *#tecnologia*, *#ai*, *#innovazione*, *#online*, *#robot*, *#google*, *#facebook*, *#web*, *#smartworking*.[25] A special attention was reserved for the topic of online privacy; in particular, with the release of 'Immuni' app for contact tracing, a lot of debates emerged about the use of private data by the government. We found a sub-community with hashtags like *#contacttracing*, *#cybersecurity*, *#cybercrime*, *#dataprotection*, *#privacy*, *#security*, *#garanteprivacy*.[26]

Green   This community contains 318 elements. In this case the main topic is sports, and, in particular, football. In fact, the most connected vertex is the hashtag

---

[18]'Basilicata', 'Calabria governed by Lega Nord', 'Sardinia' and 'Lombardy'.

[19]'Lombardy', the petition described above, assessor for welfare in Lombardy 'Giulio Gallera', president of Lombardy 'Attilio Fontana', 'Fontana resign', 'Nursing home', 'Milan'.

[20]'France', 'China', 'USA', 'Japan', 'London', 'United Kingdom', 'New York', 'Australia', 'Russia', the ex president of the United States Donald 'Trump', the president of the Spanish government Pedro 'Sanchez'.

[21]"cases", "controls"deaths", "hospitalizations", "drop", "positives".

[22]'moving', 'closed', 'municipality', 'measures', 'shops', 'order', 'mayor', 'sanctions'.

[23]'Economics', 'markets', 'companies', 'credit', 'financing', 'businesses', financial 'liquidity' and 'liquidity decree'.

[24]'Job, 'workers', 'labor unions', 'pensions', 'enterprises'.

[25]'Digital', 'technology', 'Artificial Intelligence', 'innovation', 'online', 'robot', Google', 'Facebook', 'Web', 'remote working' (during the epidemic many people worked remotely from their homes).

[26]'Guarantor of privacy'.

#spadafora (sports minister) and other hashtags with a high degree are *#atalanta*, *#bundesliga*, *#campionato*, *#dybala* (he was affected by Coronavirus), *#fifa*, *#fiorentina*, *#galliani*, *#juventus*, *#lazio*, *#liverpool*, *#seriea*[27] and many others. During the lockdown it has been discussed for a long time the problem of the restart of the European football leagues. There are also references to other sports: *#basket*, *#ciclismo*, *#djokovic*, *#pallavolo*[28] and many others.

Appendix D: Dynamics of verified users over the semantic networks

We combined the information on the communities of hashtags in the semantic network with the discursive communities of verified users described in the first section, in order to highlight the activity of verified users in generating new topics. In Fig. 11 there are the numbers of times a hashtag, belonging to one of the six communities described before, has been shared by a verified user of the 6 discursive communities identified in the Sect. 2.1. The hashtags shared by the smallest groups like PD, IV and FI are more distributed among the communities than those by DX, M5S and MEDIA communities. In detail, DX's members shared mostly hashtag from the blue and the red community, focusing on politics

Figure 11 *In the graphics there are the numbers of times a hashtag, belonging to one of the semantic communities described before, has been shared by a verified user of the discursive communities*
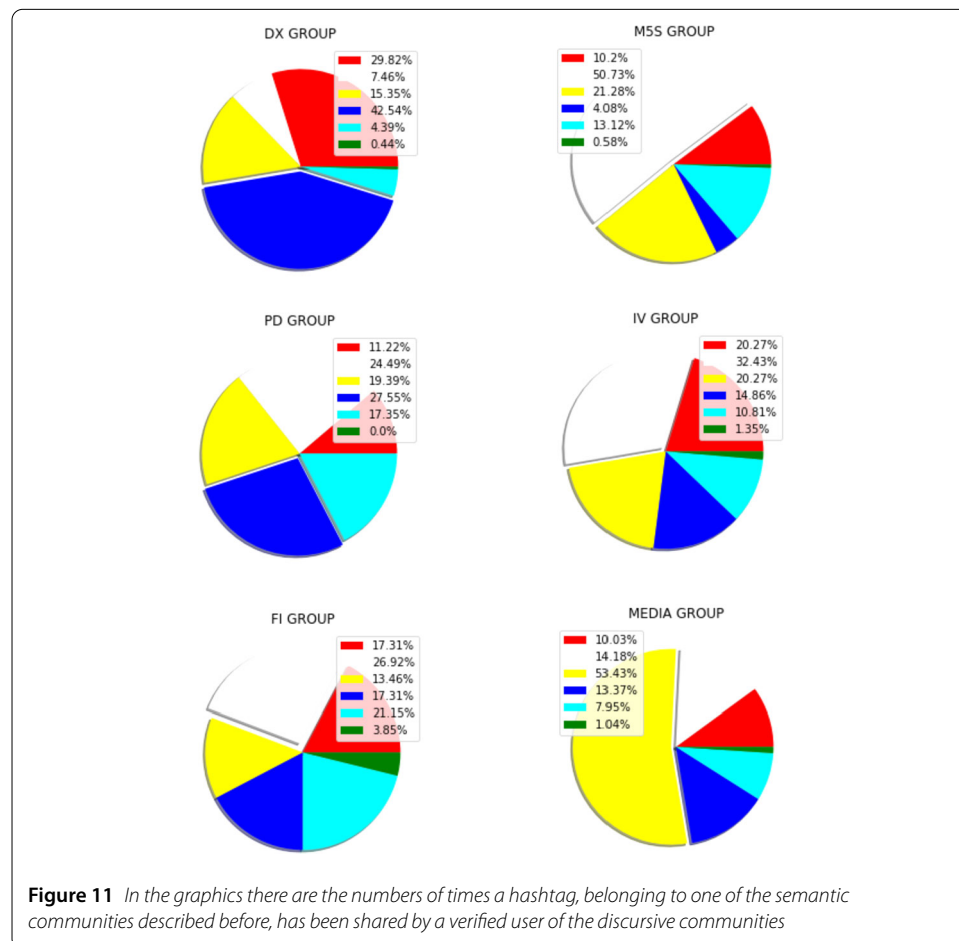
[27]'Atalanta F.C.', 'Bundesliga', 'championship', 'Paulo Dybala', 'FIFA', 'Fiorentina F.C.', 'Adriano Galliani', 'Juventus F.C.', 'Lazio F.C.', 'Liverpool F.C.', 'Serie A'.

[28]'basketball', 'cycling', 'Novak Djokovic', 'volleyball'.

and economics. The verified accounts of M5S shared contents mainly from violet community, which is about more general topics, whereas MEDIA group from the yellow one, which refers also to the international scenario and to the situations in the different Italian regions.

Author details
[1] University of Salento, P.zza Tancredi 7, 73100 Lecce, Italy. [2] "Ca' Foscari" University of Venice, Dorsoduro 3246, 30123 Venice, Italy. [3] IMT School for Advanced Studies, P.zza S. Ponziano 6, 55100 Lucca, Italy. [4] Institute for Applied Computing "Mauro Picone" (IAC), National Research Council, via dei Taurini 19, 00185 Rome, Italy.

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

References
1. Rovetta A, Bhagavathula AS (2020) COVID-19-related web search behaviors and infodemic attitudes in Italy: infodemiological study. J Med Internet Res. https://doi.org/10.2196/19374
2. Celestini A, Di Giovanni M, Guarino S, Pierri F (2020) Information disorders on Italian Facebook during COVID-19 infodemic. arXiv:2007.11302
3. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M (2020) Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Nat Hum Behav 4(12):1285–1293. https://doi.org/10.1038/s41562-020-00994-6. arXiv:2004.03997
4. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, Zola P, Zollo F, Scala A (2020) The COVID-19 social media infodemic. Sci Rep 10(1):16598. https://doi.org/10.1038/s41598-020-73510-5. arXiv:2003.05004
5. Yang K-C, Pierri F, Hui P-M, Axelrod D, Torres-Lugo C, Bryden J, Menczer F (2020) The COVID-19 infodemic: Twitter versus Facebook. arXiv:2012.09353
6. Caldarelli G, Nicola RD, Petrocchi M, Pratelli M, Saracco F (2021) Flow of online misinformation during the peak of the COVID-19 pandemic in Italy. EPJ Data Sci 10:34. https://doi.org/10.1140/EPJDS/S13688-021-00289-4
7. Patuelli A, Caldarelli G, Lattanzi N, Saracco F (2021) Firms' challenges and social responsibilities during COVID-19: a Twitter analysis. arXiv:2103.06705
8. Joint communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Tackling COVID-19 Disinformation—Getting the facts right (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020JC0008&from=EN
9. González-Bailón S, Borge-Holthoefer J, Moreno Y (2013) Broadcasters and hidden influentials in online protest diffusion. Am Behav Sci 57(7):943–965. https://doi.org/10.1177/0002764213479371
10. Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M (2015) Fame for sale: efficient detection of fake Twitter followers. Decis Support Syst 80:56–71
11. Stella M, Cristoforetti M, De Domenico M (2019) Influence of augmented humans in online interactions during voting events. PLoS ONE 14(5):e0214210. https://doi.org/10.1371/journal.pone.0214210
12. Ciampaglia GL, Nematzadeh A, Menczer F, Flammini A (2018) How algorithmic popularity bias hinders or promotes quality. Sci Rep. https://doi.org/10.1038/s41598-018-34203-2. arXiv:1707.00574
13. Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. Commun ACM 59(7):96–104
14. Yang K, Varol O, Davis CA, Ferrara E, Flammini A, Menczer F (2019) Arming the public with AI to counter social bots. CoRR. arXiv:1901.00912
15. Cresci S, Petrocchi M, Spognardi A, Tognazzi S (2019) Better safe than sorry: an adversarial approach to improve social bot detection. In: 11th international ACM web science conference, pp 47–56

16. Bovet A, Makse HA (2019) Influence of fake news in Twitter during the 2016 US presidential election. Nat Commun 10(1):7
17. Becatti C, Caldarelli G, Lambiotte R, Saracco F (2019) Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections. Palgrave Commun 5:91
18. Caldarelli G, De Nicola R, Del Vigna F, Petrocchi M, Saracco F (2020) The role of bot squads in the political propaganda on Twitter. Commun Phys 3(1):81. https://doi.org/10.1038/s42005-020-0340-4. arXiv:1905.12687
19. Radicioni T, Pavan E, Squartini T, Saracco F (2020) Analysing Twitter semantic networks: the case of 2018 Italian elections. arXiv:2009.02960
20. Radicioni T, Squartini T, Pavan E, Saracco F (2021) Networked partisanship and framing: a socio-semantic network analysis of the italian debate on migration. arXiv:2103.04653
21. Squartini T, Garlaschelli D (2017) Maximum-entropy networks. Pattern detection, network reconstruction and graph combinatorics. Springer, Berlin, p 116
22. Cimini G, Squartini T, Saracco F, Garlaschelli D, Gabrielli A, Caldarelli G (2018) The statistical physics of real-world networks. Nat Rev Phys 1(1):58–71. https://doi.org/10.1038/s42254-018-0002-6
23. Squartini T, Garlaschelli D (2011) Analytical maximum-likelihood method to detect patterns in real networks. New J Phys 13:083001. https://doi.org/10.1088/1367-2630/13/8/083001
24. Mastrandrea R, Squartini T, Fagiolo G, Garlaschelli D (2014) Enhanced reconstruction of weighted networks from strengths and degrees. New J Phys 16(4):043022
25. Saracco F, Di Clemente R, Gabrielli A, Squartini T (2015) Randomizing bipartite networks: the case of the World Trade Web. Sci Rep 5:10595
26. Saracco F, Di Clemente R, Gabrielli A, Squartini T (2016) Detecting early signs of the 2007–2008 crisis in the world trade. Sci Rep 6(1):30286. https://doi.org/10.1038/srep30286
27. Pugliese E, Cimini G, Patelli A, Zaccaria A, Pietronero L, Gabrielli A (2019) Unfolding the innovation system for the development of countries: coevolution of science, technology and production. Sci Rep 9(1):16440. https://doi.org/10.1038/s41598-019-52767-5
28. Squartini T, van Lelyveld I, Garlaschelli D (2013) Early-warning signals of topological collapse in interbank networks. Sci Rep 3(1):3357
29. Gualdi S, Cimini G, Primicerio K, Di Clemente R, Challet D (2016) Statistically validated network of portfolio overlaps and systemic risk. Sci Rep 6:39467
30. Squartini T, Caldarelli G, Cimini G, Gabrielli A, Garlaschelli D (2018) Reconstruction methods for networks: the case of economic and financial systems. Phys Rep 757:1–47. https://doi.org/10.1016/j.physrep.2018.06.008
31. Becatti C, Caldarelli G, Saracco F (2019) Entropy-based randomization of rating networks. Phys Rev E 99(2):022306
32. Saracco F, Straka MJ, Di Clemente R, Gabrielli A, Caldarelli G, Squartini T (2017) Inferring monopartite projections of bipartite networks: an entropy-based approach. New J Phys 19(5):16. https://doi.org/10.1088/1367-2630/aa6b38
33. Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106(4):620–630
34. Marchese E, Caldarelli G, Squartini T (2021) Detecting mesoscale structures by surprise. arXiv:2106.05055
35. Gemmetto V, Cardillo A, Garlaschelli D (2021) Irreducible network backbones: unbiased graph filtering via maximum entropy. arXiv:1706.00230
36. Squartini T, Caldarelli G, Cimini G, Gabrielli A, Garlaschelli D (2018) Reconstruction methods for networks: the case of economic and financial systems. Phys Rep 757:1–47
37. Mora T, Walczak A, Bialek W, Callan CG Jr (2010) Maximum entropy models for antibody diversity. Proc Natl Acad Sci USA 12(107):5405–5410. https://doi.org/10.1073/pnas.1001705107
38. Bialek W, Cavagna A, Giardina I, Mora T, Silvestri E, Viale M, Walczak AM (2012) Statistical mechanics for natural flocks of birds. Proc Natl Acad Sci USA 13(109):4786–4791. https://doi.org/10.1073/pnas.1118633109
39. Mora T, Walczak AM, Del Castello L, Ginelli F, Melillo S, Parisi L, Viale M, Cavagna A, Giardina I (2016) Local equilibrium in bird flocks. Nat Phys 12:1153–1157. https://doi.org/10.1038/nphys3846
40. Lezon TR, Banavar JR, Cieplak M, Maritan A, Fedoroff NV (2006) Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. Proc Natl Acad Sci USA 50(103):19033–19038. https://doi.org/10.1073/pnas.0609152103
41. Monechi B, Ibáñez-Berganza M, Loreto V (2020) Hamiltonian modelling of macro-economic urban dynamics. R Soc Open Sci 7:200667. https://doi.org/10.1098/rsos.200667
42. Cinelli M, Morales GDF, Galeazzi A, Quattrociocchi W, Starnini M (2021) The echo chamber effect on social media. Proc Natl Acad Sci USA 118:e2023301118. https://doi.org/10.1073/PNAS.2023301118
43. Conover M, Ratkiewicz J, Francisco M (2011) Political polarization on Twitter. Icwsm. https://doi.org/10.1021/ja202932e
44. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008
45. Chung F, Lu L (2002) Connected components in random graphs with given expected degree sequences. Ann Comb 6:125–145. https://doi.org/10.1007/PL00012580
46. Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. Phys Rev E, Stat Nonlinear Soft Matter Phys. https://doi.org/10.1103/PhysRevE.76.036106. arXiv:0709.2938
47. Adamic LA, Glance NS (2005) The political blogosphere and the 2004 U.S. election: divided they blog. In: 3rd international workshop on link discovery, LinkKDD 2005, Chicago, Illinois, USA, August 21–25, pp 36–43
48. Del Vicario M, Vivaldo G, Bessi A, Zollo F, Scala A, Caldarelli G, Quattrociocchi W (2016) Echo chambers: emotional contagion and group polarization on Facebook. Sci Rep. https://doi.org/10.1038/srep37825
49. Zollo F, Bessi A, Del Vicario M, Scala A, Caldarelli G, Shekhtman L, Havlin S, Quattrociocchi W (2017) Debunking in a world of tribes. PLoS ONE 12(7):e0181821. https://doi.org/10.1371/journal.pone.0181821
50. Bradshaw S, Howard P (2017) The global organization of social media disinformation campaigns. J Int Affairs 71(1.5)
51. Bradshaw S, Howard P (2018) How does junk news spread so quickly across social media? Algorithms, advertising and exposure in public life. Oxford Internet Institute—White paper
52. Park J, Newman MEJ (2004) Statistical mechanics of networks. Phys Rev E 70(6):66117. https://doi.org/10.1103/PhysRevE.70.066117

53. Garlaschelli D, Loffredo MI (2008) Maximum likelihood: extracting unbiased information from complex networks. Phys Rev E, Stat Nonlinear Soft Matter Phys 78(1):015101. https://doi.org/10.1103/PhysRevE.78.015101
54. Strona G, Nappo D, Boccacci F, Fattorini S, San-Miguel-Ayanz J (2014) A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. Nat Commun 5:4114
55. Carstens CJ (2015) Proof of uniform sampling of binary matrices with fixed row sums and column sums for the fast curveball algorithm. Phys Rev E, Stat Nonlinear Soft Matter Phys 91(4):042812. https://doi.org/10.1103/PhysRevE.91.042812
56. Squartini T, Mol JD, Hollander FD, Garlaschelli D (2015) Breaking of ensemble equivalence in networks. Phys Rev Lett. https://doi.org/10.1103/PhysRevLett.115.268701
57. Squartini T, Mastrandrea R, Garlaschelli D (2015) Unbiased sampling of network ensembles. New J Phys 17:023052. https://doi.org/10.1088/1367-2630/17/2/023052
58. Bruno M, Saracco F, Garlaschelli D, Tessone CJ, Caldarelli G (2020) The ambiguity of nestedness under soft and hard constraints. Sci Rep 10:19903. https://doi.org/10.1038/s41598-020-76300-1
59. Simmons BI, Sweering MJM, Schillinger M, Dicks LV, Sutherland WJ, Clemente RD (2019) bmotif: a package for motif analyses of bipartite networks. Methods Ecol Evol 10:695–701. https://doi.org/10.1111/2041-210X.13149
60. Hong Y (2013) On computing the distribution function for the Poisson binomial distribution. Comput Stat Data Anal 59(1):41–51
61. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57(1):289–300
62. Newman M (2010) Networks: an introduction. Oxford University Press, New York
63. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99(12):7821–7826. https://doi.org/10.1073/pnas.122653799. arXiv:cond-mat/0112110
64. Vallarano N, Bruno M, Marchese E, Trapani G, Saracco F, Squartini T, Cimini G, Zanon M (2021) Fast and scalable likelihood maximization for exponential random graph models. arXiv:2101.12625
65. Yang C, Harkreader R, Gu G (2013) Empirical evaluation and new design for fighting evolving Twitter spammers. IEEE Trans Inf Forensics Secur 8(8):1280–1293
66. Gentzkow MA, Shapiro JM (2010) Ideological segregation online and offline. SSRN Electron J. https://doi.org/10.2139/ssrn.1588920
67. Blaschke T, Hay GJ, Weng Q, Resch B (2011) Collective sensing: integrating geospatial technologies to understand urban systems—an overview. Remote Sens 3:1743–1776. https://doi.org/10.3390/rs3081743
68. Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. Science 348:1130–1132. https://doi.org/10.1126/science.aaa1160
69. Bastos M, Mercea D, Baronchelli A (2018) The geographic embedding of online echo chambers: evidence from the brexit campaign. PLoS ONE 13:e0206841. https://doi.org/10.1371/journal.pone.0206841
70. Dong X, Morales AJ, Jahani E, Moro E, Lepri B, Bozkaya B, Sarraute C, Bar-Yam Y, Pentland A (2020) Segregated interactions in urban and online space. EPJ Data Sci 9:20. https://doi.org/10.1140/epjds/s13688-020-00238-7
71. Tóth G, Wachs J, Clemente RD, Jakobi Á, Ságvári B, Kertész J, Lengyel B (2021) Inequality is rising where social network segregation interacts with urban topology. Nat Commun 12:1143. https://doi.org/10.1038/s41467-021-21465-0