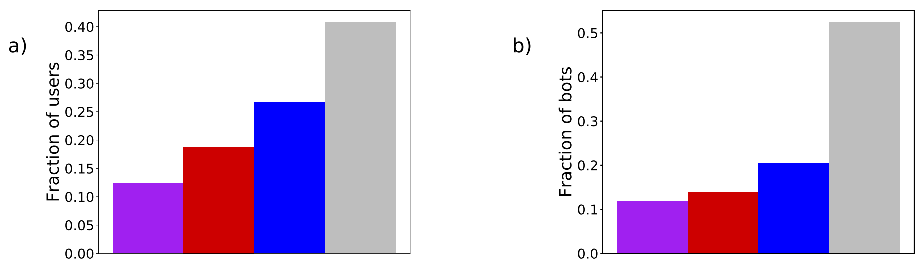
Supplementary Information

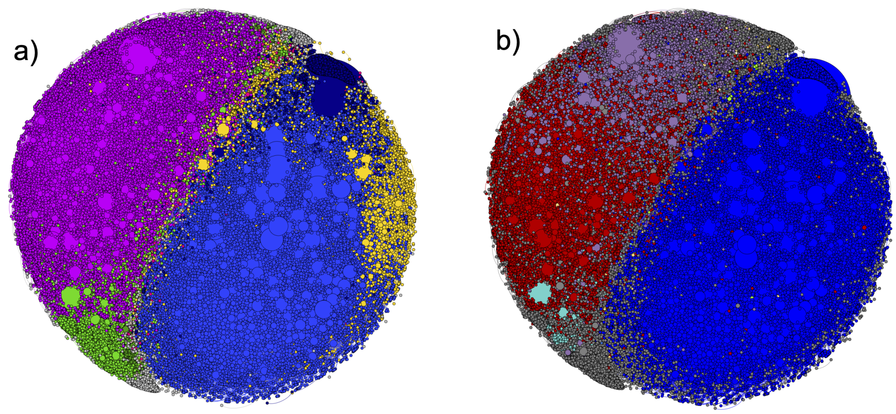
March 17, 2020



Supplementary Figure 1: **Users and bots affiliated to the different communities, after the contagion of polarization.** Panel **a)**: fraction of users affiliated to the different communities, after the contagion of polarization. Nodes are eggplant purple for NGO's, media accounts and left wing politicians, tomato red for the Democratic Party community and Mediterranean blue for the pro-government users. In gray, users which cannot be assigned uniquely to one community. Panel **b)**: the same histogram of panel **a)** for bots. Percentages of non-gray groups are lower than the analogous ones for all users, but for the eggplant purple community. The (relative) extra density of bots in the purple community may be due to the high presence of automated news spreaders, commonly employed to automatically posting links to newspapers articles.

Supplementary Note 1: The Italian political situation at the time of the study The 2018 general elections resulted in a hung parliament, since no political party won an outright majority. After several negotiations, on May 31, two Italian parties, Lega and Movimento 5 Stelle (hereafter M5S) reached an agreement to form a new government.

Lega is a right nationalist and sovereign party founded in Italy in 1991 which has been strongly advocating to close the border to the immigrants from Northern Africa [1]. After the 2018 national election, since no coalition had the majority in the parliament, Lega decided to break its electoral alliance with Forza Italia and Fratelli d'Italia to form a government with M5S and its federal secretary, Matteo Salvini, was nominated Minister of Internal Affairs. During



Supplementary Figure 2: **Undirected network of retweet interactions**. Panel **a)**: nodes have been coloured according to memberships found by the Louvain community detection on the network. Panel **b)**: nodes coloured according to the iterative polarization assignment.

his mandate Salvini introduced strict immigration policies and repeatedly accused the NGOs operating in the Mediterranean Sea to collaborate with humans' smugglers from Lybia.

M5S is a political party founded in Italy in 2009 by the comedian and blogger Beppe Grillo and the web strategist Gianroberto Casaleggio. The movement has always refused any possible positioning in the parliament. It was variously considered populist, anti-establishment, and Eurosceptic [2, 3]. During the government with Lega, Luigi Di Maio, the Political Head of the M5S, was Minister of Economic Development, Labour and Social Policies. The M5S supported the actions of the Minister of Internal Affairs, as part of the government at that time.

The Italian Democratic Party (hereafter PD) is a social-democratic political party [4], founded in 2007. The former Prime Ministers Matteo Renzi and Paolo Gentiloni have been two of its leaders. During the period of data collection, the PD was opponent of the government and took strong positions against the immigration policies adopted by the Minister Salvini.

Forza Italia is a centre-right [5] political party in Italy, founded in 1993; its leader Silvio Berlusconi has been four times Prime Minister. During the period of data collection, Forza Italia was opponent of the government, but was in relative agreement on its immigration policies.

Finally, Fratelli d'Italia is a national-conservative political party in Italy, founded in 2012 and whose leader is Giorgia Meloni (`http://www.parties-and-elections.eu/italy.html`). In February, 2019, Fratelli d'Italia was part of the opposition; nevertheless, on the immigration politics, Fratelli d'Italia had a position very close to the one of Lega.

The clusters identified in Figure 2 of the main text consist of:

Mediterranean blue cluster Among others, in this set we find the Minister of Internal Affairs, and Deputy Prime Minister, Matteo Salvini, the Minister of Infrastructures and Transports Danilo Toninelli, the Minister of Health Giulia Grillo; the right wing verified accounts including the national-conservative party Fratelli d'Italia and its leader Giorgia Meloni; the Forza Italia party and some of its representatives, as Giovanni Toti; Lega with other politicians than the above mentioned Salvini; M5S representatives, like Carlo Sibilia and Roberta Lombardi.

Tomato red cluster Among others, in this set we find the candidates to the PD primary elections Nicola Zingaretti, Maurizio Martina and Roberto Giachetti, its previous leader Matteo Renzi and the previous Prime Minister Paolo Gentiloni; Pierluigi Bersani (former secretary of PD), Laura Boldrini (former President of the Chamber of Deputies) and Marco Furfaro (a left-wing politician): all the latters, at the time of data collection, belonged to parties on the left of PD.

Eggplant purple cluster We find in this set: the NGOs, as the Italian chapters of ActionAid, Medecins Sans Frontieres, Unicef, Amnesty International, and OpenArms; Giuseppe Civati (a former deputy of PD), Nichi Vendola (the former governor of the Apulia region), Luigi De Magistris (the major of Naples) and Enrico Rossi (the actual governor of Tuscany; all of them are considered as lefter than PD); the online newspaper Repubblica, il Post, Internazionale, la Valigia Blu, Rolling Stone Italia and their journalists; many public figures as the film directors Gabriele Muccino and Ferzan Ozpetek, the actors Alessandro Gassmann and Pif, the singers Fiorella Mannoia, Nina Zilli, Frankie Hi-NRG.

Orange cluster The newspaper "l'Osservatore Romano" and the website "Vatican News" and "Vatican Radio" are part of this group.

Supplementary Note 2: The Bipartite Configuration Model In the present section, we outline the procedure to obtain the probability per link of the Bipartite Configuration Model (*BiCM*, [6]), i.e., the extension of the entropy-based null-models to bipartite networks [7, 8].

Let us start with a real bipartite network in which we label the two layers with L and Γ. We will refer to N_{L} and N_{Γ}, respectively, for the dimensions of the two layers; nodes on the two layers are indicated respectively with i and

α. A bipartite undirected network can be described in terms of a biadjacency matrix, i.e., a $N_{\mathrm{L}} \times N_{\Gamma}$ matrix $\mathbf{M}$ whose generic entry $m_{i\alpha}$ is 1 if a link connects the node $i \in \mathrm{L}$ to the node $\alpha \in \Gamma$ and 0 otherwise. The degree of a generic node $i \in \mathrm{L}$, i.e., the number of links of i, can be expressed as $k_i = \sum_{\alpha \in \Gamma} m_{i\alpha}$; the analogous for a generic $\alpha \in \Gamma$ is $k_\alpha = \sum_{i \in \mathrm{L}} m_{i\alpha}$.

Let us define $\mathcal{G}_{\mathrm{Bi}}$ the *ensemble* of all possible graphs with the same number of nodes per layer as in the real network. We label with an asterisk $*$ all the quantities of the real network, but for N_{L} and N_{Γ}, since they are going to be fixed on the entire ensemble. If every graph $G_{\mathrm{Bi}} \in \mathcal{G}_{\mathrm{Bi}}$ is equipped with a probability $P(G_{\mathrm{Bi}})$, we can define the Shannon entropy for the ensemble as

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

S represents the uncertainty we have over the system and we aim at maximising it in order to be as general as possible. However, we discount some information about the system, thus maximising our uncertainty but for some quantities that we want to conserve over the ensemble. This goal can be obtained by maximising the Shannon entropy, constrained such that the average values of some quantities of interest are fixed. If the constraints are expressed as a vector $\vec{C}(G_{\mathrm{Bi}})$ it can be shown that the probability per graph reads [9]:

$$P(G_{\mathrm{Bi}}) \sim e^{-\vec{\theta} \cdot \vec{C}(G_{\mathrm{Bi}})},$$

where $\vec{C}(G_{\mathrm{Bi}})$ is the vector of the constraints evaluated on the graph G_{Bi} and $\vec{\theta}$ is the vector of the relative Lagrangian multipliers [9]. If we constrain the value of the degree sequence, i.e., all k_i's and k_α's, it can be shown that the probability per graph factorises in independent probabilities per link:

$$P(G_{\mathrm{Bi}}|\vec{x}, \vec{y}) = \prod_{i \in \mathrm{L}} \prod_{\alpha \in \Gamma} p_{i\alpha}^{m_{i\alpha}} (1 - p_{i\alpha})^{1 - m_{i\alpha}},$$

where

$$p_{i\alpha} = \frac{x_i y_\alpha}{1 + x_i y_\alpha}. \tag{1}$$

If $\vec{\theta}$ and $\vec{\eta}$ are the Lagrangian multipliers of the degree sequences of the L and Γ layers, $x_i = e^{-\theta_i}$ and $y_\alpha = e^{-\eta_\alpha}$. x_i and y_α are called *fitnesses*. Fitnesses encode the attitude, respectively of nodes i and α, to establish links [10].

Equation (1) is formal: so far, we just impose that, over the ensemble, the average of the degree sequence must be fixed. In order to 'tailor' the ensemble on the real network, we have to impose that the average degree sequence is the one measured on the real network [11, 12], i.e.,

$$\begin{cases} \langle k_i \rangle = \sum_\alpha p_{i\alpha} = k_i^* \\ \langle k_\alpha \rangle = \sum_i p_{i\alpha} = k_\alpha^* \end{cases}. \tag{2}$$

In this way, we have an ensemble of networks to compare our results with, showing exactly the same degree sequence, but being as general as possible regarding all the other features.

The solution of System (2) can only be numerical, since no analytic form can be found. Due to the dimensions of the system under analysis, we approximate the probability per link as $p_{i\alpha} \simeq x_i y_\alpha$; the approximation is justified as long as the network is sparse, in particular when $x_i, y_\alpha \ll 1$. In such a case, the null model is analogous to the one of [13] and the solution of System (2) is analytic:

$$p_{i\alpha}^{\text{CLA}} = \frac{k_i^* k_\alpha^*}{m^*}, \tag{3}$$

where m^* is the total number of links measured on the real network and "CLA" stays for "Chung-Lu Approximation".

Supplementary Note 3: Validated Monopartite Projection We can make use of the *BiCM* in order to project the information contained in the original network, discounted by the degree sequence contribution. The main idea was developed in [14]: due to the independence of the probabilities per link, the probability that both nodes $i, j \in \mathrm{L}$ are connected to node $\alpha \in \Gamma$ is

$$P(V_\alpha^{ij}) = p_{i\alpha} p_{j\alpha},$$

where V_α^{ij} is the event of i and j linking α: such a pattern was described in [6] as a V-motif since, if the two layers are represented as two horizontal lines of vertices, it draws a "V" between the layers. More formally, they are $K_{2,1}$ stars [15]. If we are interested in how much similar two nodes are in terms of common connections, we can compare the number of common links in the real network $(V^{ij})^* = \sum_{\alpha \in \Gamma} (V_\alpha^{ij})^*$ with its theoretical distribution, according to the *BiCM*. Since all the probabilities $P(V_\alpha^{ij})$ are different -in general- the theoretical probability distribution is a Poisson-binomial [16–18], i.e., the extension of a Binomial distribution in which every Bernoulli event has a different probability. If the probabilities per $V-$motif are particularly low (as it is the case of the present study), the Poisson-binomial distribution can be approximated with a Poisson distribution [18], in which the parameter is

$$\lambda = \langle V^{ij} \rangle = \sum_{\alpha \in \Gamma} p_{i\alpha} p_{j\alpha} \overset{\text{CLA}}{=} \frac{k_i^* k_j^*}{(m^*)^2} \sum_{\alpha \in \Gamma} (k_\alpha^*)^2,$$

where in the last step we implement the Chung-Lu approximation (3).

Once we have all the theoretical distributions, we can assign to every V^{ij} a p-value and then validate all the statistically significant ones. In the present article, we set the statistical significant level $\alpha = 0.01$ and implement the FDR (*False Discovery Rate*, [19]) for the validation. FDR is one of the most effective methods for multiple testing hypothesis, i.e., to claim the statistical significance of many p-values at the same time, since it permits to control the number of False Positives [19]. The procedure is pretty simple: once all N p-values are ordered from the smallest to the greatest, i.e.,

$$\text{p-value}_1 \leq \text{p-value}_2 \leq \cdots \leq \text{p-value}_N,$$

the effective statistical significance threshold is the greatest $i\frac{\alpha}{N}$ satisfying

$$\text{p-value}_i \leq i\frac{\alpha}{N}.$$

Supplementary Note 4: Iterative Procedure for Assigning a Political Inclination to Unverified Users As mentioned in the main text, just almost one half of the unverified users interact with verified ones. Since this interaction is used to infer the inclination of users towards a political point of view directly from data, we iteratively extend the procedure for the inclination assignment. The idea is to consider all the unpolarized users and assign them the inclination of the majority of their followers and followees. If the polarization index ρ (as defined in (1) of the main text) of a given unpolarized node is lower than 0.5, then the majority of its neighbours is non polarized. Thus, the node re-enters the polarization step, until no polarization can be assigned anymore. Noticeably, the previous threshold -0.5- is particularly high, since we have identified 3 big communities. The nodes resulting in grey in the first bipartite validation have been considered in the contagion polarization too.

The algorithm stops after 10 steps, and we were able to assign a political affiliation to almost the 27% of the unpolarized users, resulting in the 15% of the entire set of accounts. Even if the previous percentage seems quite small, it is not when we focus on the nodes in the directed validated projection, as it can be seen in Figure 11 of the main text. The amount of unpolarized users represents the 8.3% of the total number of validated accounts. Literally, the polarization by contagion was able to assign a polarization to almost the 58% of unpolarized nodes in the directed projection.

Supplementary Note 5: The Bipartite Directed Configuration Model As a side product of defining a Monopartite Configuration Model with a community structure, in [20] the directed extension of the *BiCM* was presented; hereafter, we are going to call it Bipartite Directed Configuration Model (*BiDCM*). In the following, we will use the formalism of [21], since it is tailored on our application. Thus, we will indicate the two layers as U (i.e., users) and P (i.e., posts), and N_U and N_P are, respectively, their dimensions. A bipartite directed network can be described by using two $N_U \times N_P$ biadjacency matrices, respectively $\mathbf{T}$ (i.e., the tweets), indicating links going from U to P, and $\mathbf{R}$ (i.e., retweets), indicating links from P to U. Considering the exchange of content in the original network, we use the $\mathbf{T}$ matrix to describe the authorship of tweets, while $\mathbf{R}$ is for retweets. If we focus on a generic user u, her/his out- and in-degrees are respectively

$$\kappa_u^{\text{out}} = \sum_{p \in P} t_{up}; \qquad \kappa_u^{\text{in}} = \sum_{p \in P} r_{up}$$

The two quantities represent, respectively, the number of tweets the user u wrote and the number of tweets u retwetted. Analogously, for the generic post p is

$$\kappa_p^{\text{in}} = \sum_{u \in U} t_{up} = 1; \qquad \kappa_p^{\text{out}} = \sum_{u \in U} r_{up}.$$

As it can be seen in the previous definitions, we have a great simplification, due to the nature of our system: since there can only be a single author for a given post, the in-degree of a post is always 1, for every p. This feature is going to introduce a great simplification in the following calculations.

By following the same track of the previous section, if we call the ensemble of bipartite directed graphs $\mathcal{G}_{\text{BiD}}$, by maximising the entropy of the system constraining the entire (directed) degree sequence, it can be shown that the probability for a generic bipartite directed graph $G_{\text{BiD}} \in \mathcal{G}_{\text{BiD}}$ still factorises in terms of probabilities per link [20, 21]:

$$P(G_{\text{BiD}}|\vec{z}, \vec{z}', \vec{\zeta}, \vec{\zeta}') = \prod_{u \in U} \prod_{p \in P} q_{up}^{t_{up}} (1 - q_{up})^{1-t_{up}} \cdot \prod_{u' \in U} \prod_{p' \in P} (q_{u'p'})^{r_{u'p'}} (1 - q_{u'p'})^{1-r_{u'p'}},$$

where

$$q_{up} = \frac{z_u \zeta_p}{1 + z_u \zeta_p}; \qquad q'_{u'p'} = \frac{z'_{u'} \zeta'_{p'}}{1 + z'_{u'} \zeta'_{p'}} \tag{4}$$

are respectively the probabilities of u being an author of the post p and of u' being a retwitter of the post p'. Again, by maximising the likelihood of the real systems, we have that, among the other conditions, $\langle \kappa_p^{\text{in}} \rangle = 1 \, \forall p \in P$. The previous condition implies that all the fitnesses ζ are equal for every node in P and it can be shown that $q_{up} = \frac{(\kappa_u^{\text{out}})^*}{N_P} \, \forall p \in P$ [21].

Supplementary Note 6: Directed Validated Projection In order to extract those relationships among users that cannot be simply referred to their activity and the virality of posts, we project the directed bipartite network on the layer of users, following the same strategy of the validated projection for undirected bipartite networks. Differently from the previous validated projection, in the present case the final result is going to be a directed monopartite network in which links flow from the significant source of post to the significant retwetters.

In the present study, the approach is analogous to the one in [21], but for two crucial differences. First, in [21] we analysed the contribution of retwetted posts only, while here we also consider the contribution of non-retwetted ones. This choice affects only the probabilities q's: the rationale is that we aim at detecting if an account retweets all the posts of a certain user or it makes a selection on the messages of the given author. Secondly, we not only limit the investigation to the content flow, but also we explore the contribution of bot accounts to that flow. In terms of message spreading, and linked to our exploration on the activities, retwetting all the posts of a certain user is probably more critical than selecting part of them.

We call $\mathcal{V}_p^{uv}$ the event of v retwetting the post p written by u: according to the *BiDCM*, its probability is simply

$$P(\mathcal{V}_p^{uv}) = q_{up}q'_{vp} = \frac{(\kappa_u^{\text{out}})^*}{N_P}q'_{vp} \tag{5}$$

As in the case of the undirected bipartite network, due to the dimension of the system, we approximate the Poisson-binomial distribution of $\mathcal{V}^{uv}$ with a Poisson distribution in which the parameter is

$$\lambda = \langle \mathcal{V}^{uv} \rangle = \sum_{p \in P} P(\mathcal{V}_p^{uv}) = \sum_{p \in P} \frac{(\kappa_u^{\text{out}})^*}{N_P} q'_{vp} = \frac{(\kappa_u^{\text{out}})^*}{N_P} \sum_{p \in P} q'_{vp} = \frac{(\kappa_u^{\text{out}})^*(\kappa_v^{\text{in}})^*}{N_P},$$

where, in the last steps, we made use of equation (5). It is worth noting that, with respect to the previous case, in the actual projected validation we do not use the Chung-Lu approximation in order to simplify the calculations, but we take advantage of the fact that any post has just one author.

Finally, we can associate to each observed $\mathcal{V}^{uv}$ a p-value via the previous distributions and we validate them via the above mentioned FDR [19]; in the present case too, the statistical significant level is set to $\alpha = 0.01$.

Supplementary Note 7: Bot Squads The analysis of the role of bots in the validated backbone of the traffic of messages on Twitter let us observe the presence of automated accounts that follow more than one genuine user. We call such groups 'bot squads'. Remarkably, we find two groups of genuine users sharing more than 2 automated accounts. The biggest users group includes 12 of the top 20 hubs (9 in the top 10). Referring to Figure 8 in the main text, we detect the presence of a subgraph of 22 hubs sharing 22 bots among their followers: literally, each of these 22 bots does not retweet the content of a single hub, but of more than one user in the subgraph. Thus, the bot squad increases the visibility of the common followees.

Among the hubs in the subgraph, we can find the account of the Minister of Internal Affairs Mr. Salvini, as well as the Minister of Infrastructures Danilo Toninelli. Other verified accounts are the one of Giorgia Meloni of Fratelli d'Italia, a right wing party, and the official accounts of Lega (Mr. Salvini's party). In this set, we find also the news website supported by Casa Pound, a neo-fascist Italian party, two journalists from the same website, the director of the newspaper La Verità, as well as other politicians of the Lega party.

Figure 9 in the main text shows the subgraph of the violet bot squad: there, the genuine accounts are much less efficient in delivering their messages, since the strongest hub, representing a popular newscast, ranks 176^{th} in the overall hub score. Moreover, while the blue subgraph included 178 nodes, the violet one contains just 58 accounts. In this subgraph, we find the presence of several NGOs, some NGO representatives (coming, for instance, from 'Comunità di Sant'Egidio', a Catholic NGO, and from MOAS, a NGO active in aiding migrants in the Mediterranean sea), some journalists from the TV channel La7 and small political parties belonging to the left wing.

The incidence of bots in the two subgraphs of Figures 8 and 9 in the main text is relatively similar, being the 87% for the first subgraph and 79% for the second one. Instead, the ratio between the members of the bot squads (i.e., the number of shared bots) over the total number of genuine users in the subgraphs is not: the former is exactly 1, while the latter is around 0.58. Interestingly, in both sets, the hubs rarely retweet between each other in a significant way (in fact, only 3 links can be found among them). They leave the duty of spreading the content of the partners to the bots.

Supplementary Note 8: Alternative approaches to assign a political inclination The adoption of the validated projection approach for detecting the polarization of the users may appear as unnecessarily complicated, with respect to launching a community detection algorithm on the monopartite network of interactions among all the accounts. However, the validation process removes random noise from the system, thus providing a (validated) network in which the communities are clearer. The monopartite network of undirected interactions can be found in Supplementary Figure 2: in the left panel, nodes have been colored according to the result of the Louvain community detection algorithm; in the right panel, the communities have been obtained by the (iterative) polarization described in the main text. There are some similarities between the two partitions: the blue communities almost overlap (but for some relevant differences that we are going to discuss in the following). However, on the one hand, the left partition pinpoints a yellow group (containing mostly M5S politicians and journalists); on the other hand, the right partition reveals two main groups in the upper cluster, instead of the single one on the left.

At first glance, the two panels may seem more or less equivalent, but literally they are not. If we consider the affiliation of the various verified users, the differences are somehow striking. Firstly, the pink community in the left panel contains the verified accounts of both the red and the purple community of the right panel, thus mixing the two groups. Secondly, the yellow community in the left partition reveals a group that is not detected in the right one, i.e., the one of politicians and journalists supporting the Movimento 5 Stelle (M5S) party. However, this cluster also contains other newspapers far from M5S, as 'La Stampa', 'Il Corriere della Sera', 'Il Sole24ore', a gossip website ('Dagospia') and some TV newscasts as 'SkyNews 24', 'RaiNews' and 'Tg La7'. The blue community in the left panel is quite similar to the one in the right panel, but for the presence of some Democratic Party parliamentarians and for the absence of Mr. Matteo Salvini. The latter fact is especially odd, since the two official accounts of Lega, i.e., the political party of Mr. Salvini, are still in the blue community. In the dark blue community of the left panel, absent in the right panel, we find the account of Mr. Salvini, together with some French politicians from the *Rassemblement National*, as well as some German politicians from the *Alternative für Deutschland*. At a first sight, this sounds reliable, since Lega, Rassemblement National, Alternative für Deutschland are allied in the European Parliament. Nevertheless, the presence of some parliamentarians of the

Italian Democratic Party weaken the reliability of such a cluster. Analogously, the quite big green community in the left panel contains a pot-pourri of accounts, like, e.g., the Italian embassy in Greece and the referees of a popular Italian TV dance contest, ‘Ballando con le Stelle’.

We argue that the non validated partition, e.g., the one described in the left panel, is reasonable, but the amount of noise is much higher with respect to the validated one, i.e., a relatively high number of verified nodes has a clearly wrong membership. In the validated projection, the number of evidently wrong assignments is much lower, being limited to few center-left wing journalists in the community of the right wing. Due to such a difference, in this paper we have adopted the validated projection approach.

Supplementary References

[1] Balmer, C. Italy’s League leaves northern bastions, bangs anti-migrant drum. February 2018 Online: https://www.reuters.com/article/us-italy-election-league-idUSKBN1FT1Q9.

[2] Lanzone, M. *The “post-modern” populism in Italy: The case of the Five Star Movement* (Emerald Group Publishing, 2014).

[3] Jungar, A.-C. *Repercussions of right-wing populism for European integration* (Edward Elgar Publishing, 2018).

[4] Pasquino, G. *Italy* (Palgrave Macmillan, 2016).

[5] Cotta, M. & Verzichelli, L. *Political Institutions of Italy* (Oxford University Press, 2012).

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

[7] Squartini, T. & Garlaschelli, D. *Maximum-entropy networks. Pattern detection, network reconstruction and graph combinatorics* (Springer International Publishing, 2017).

[8] Cimini, G. *et al.* The Statistical Physics of Real-World Networks. *Nat. Rev. Phys.* **1**, 58–71 (2018).

[9] Park, J. & Newman, M. E. J. Statistical mechanics of networks. *Phys. Rev. E* **70**, 66117 (2004).

[10] Caldarelli, G., Capocci, A., De Los Rios, P. & Muñoz, M. A. Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **89**, 258702 (2002).

[11] Garlaschelli, D. & Loffredo, M. I. Maximum likelihood: Extracting unbiased information from complex networks. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **78**, 1–5 (2008).

[12] Squartini, T. & Garlaschelli, D. Analytical maximum-likelihood method to detect patterns in real networks. *New J. Phys.* **13** (2011).

[13] Chung, F. & Lu, L. Connected Components in Random Graphs with Given Expected Degree Sequences. *Ann. Comb.* **6**, 125–145 (2002).

[14] Saracco, F. *et al.* Inferring monopartite projections of bipartite networks: An entropy-based approach. *New J. Phys.* **19**, 16 (2017).

[15] Diestel, R. *Graph Theory, 5th Edition* (Springer-Verlag Heidelberg, 2016).

[16] Volkova, A. Y. A Refinement of the Central Limit Theorem for Sums of Independent Random Indicators. *Theory Probab. Its Appl.* (2005).

[17] Deheuvels, P., Puri, M. L. & Ralescu, S. S. Asymptotic expansions for sums of nonidentically distributed Bernoulli random variables. *J. Multivar. Anal.* (1989).

[18] Hong, Y. On computing the distribution function for the Poisson binomial distribution. *Comput. Stat. Data Anal.* **59**, 41–51 (2013).

[19] Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

[20] van Lidth de Jeude, J., Clemente, R. D., Caldarelli, G., Saracco, F. & Squartini, T. Reconstructing mesoscale network structures. *Complexity* **2019**, 5120581:1–5120581:13 (2019).

[21] 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.* (2019).