The role of bot squads in the political propaganda on Twitter

Fabio Del Vigna
fabio.delvigna@imtlucca.it
IMT Scuola Alti Studi Lucca
Lucca, Italy

Guido Caldarelli
guido.caldarelli@imtlucca.it
IMT Scuola Alti Studi Lucca
Lucca, Italy
European Centre for Living Technology, Università di Venezia "Ca' Foscari"
Venice, Italy
Catchy srl
Rome, Italy
Istituto dei Sistemi Complessi CNR, Dip. Fisica, Università Sapienza
Rome, Italy

Rocco De Nicola
rocco.denicola@imtlucca.it
IMT Scuola Alti Studi Lucca
Lucca, Italy

Marinella Petrocchi
marinella.petrocchi@iit.cnr.it
Istituto di Informatica e Telematica, CNR
Pisa, Italy
IMT Scuola Alti Studi Lucca
Lucca, Italy

Fabio Saracco
fabio.saracco@imtlucca.it
IMT Scuola Alti Studi Lucca
Lucca, Italy

ABSTRACT

Social Media are nowadays the privileged channel for information spreading and news checking. Unexpectedly for most of the users, automated accounts, also known as social bots, contribute more and more to this process of news spreading. Beside the fruitful activities of benign bots, social platforms are unfortunately overwhelmed by malicious ones, which aim at perturbing the user base for altering the political, cultural, and economic perception of real-world facts. Using Twitter as a benchmark, we consider the traffic exchanged, over one month of observation, on a specific topic, namely the migration flux from Northern Africa to Italy. We measure the significant traffic of tweets only, by implementing an entropy-based null model that discounts the activity of users and the virality of tweets. Result show that social bots play a central role in the exchange of significant content. Indeed, not only the strongest hubs have a number of bots among their followers higher than expected, but furthermore a group of them, that can be assigned to the same political matrix, share a common set of bots as followers. The retwitting activity of such automated accounts amplifies the presence on the platform of the hubs' messages.

KEYWORDS

bot detection, complex networks analysis, online social networks, political debate, entropy-based null-models, Shannon entropy, null-models

ACM Reference Format:
Fabio Del Vigna, Guido Caldarelli, Rocco De Nicola, Marinella Petrocchi, and Fabio Saracco. 2019. The role of bot squads in the political propaganda on Twitter. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Since a decade microblogging platforms, like Twitter, have become prominent sources of information [23], catching breaking news and anticipating more traditional media like radio and television [20]. Indeed the 2018 Eurobarometer report on news consumption presents a clear increasing trend of popularity of online news sources with respect to traditional ones. Albeit this widespread favour, online media are not trusted as their offline counterparts [37]: in a survey conducted in autumn 2017, 59% of respondents said they trusted radio content, while only 20% said they trusted information available on online social networks. Even beside the perception of common users, the presence of fake contents has indeed been revealed in several research work, both at level of news *per se* and of fake accounts contributing to spreading them, see, e.g., [7, 9, 14, 27, 30]. Actually, on Twitter we assist to the proliferation of social accounts governed - completely or in part - by pieces of software that automatically create, share, and like contents on the platform. Such software, also known as *social bots* - or simply *bots* - can be programmed to automatically post information about news of any kind and even to

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

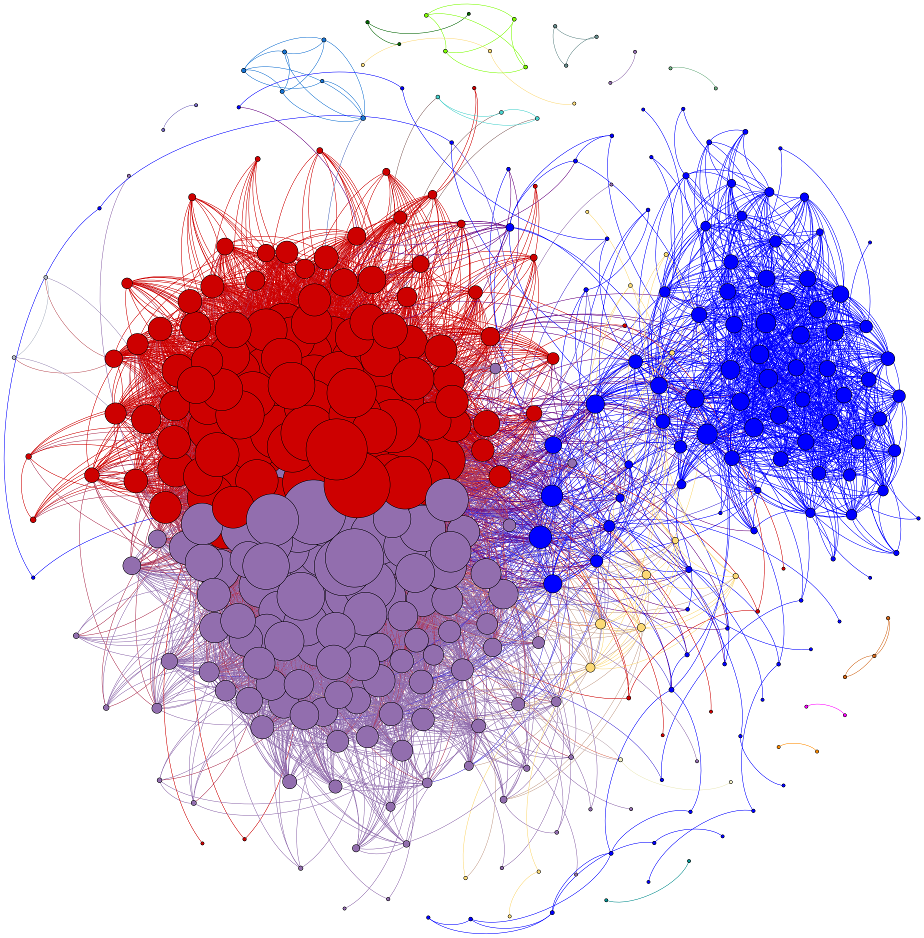

Figure 1: Affiliation network as resulted from the validated projection of the bipartite network of verified and unverified users. The 3 main communities have a clear interpretation: in blue there are the accounts in the right wing and Movimento 5 Stelle; in red the ones from the Italian Democratic Party; in violet the community of NGO's and official newspaper and newcast accounts. The dimension of the nodes is proportional to their degree.

provide help during emergencies. Unfortunately, our online ecosystem is constantly threatened by malicious social bots, recently deemed responsible for tampering with online discussions about major political election in western countries, including the 2016 US presidential elections, and the UK Brexit referendum [2, 4, 13, 15]. Recent work demonstrates that social bots are particularly active in spreading low credibility content and amplifying their visibility [30]. They also target influential people, bombarding them with hateful contents [35], and they even interact with users according to their political opinion[18]. Bots' actions do not spare financial markets: as much as 71% of the authors of suspicious tweets about US stocks have been classified as bots by a state-of-the-art spambot detection algorithm [10].

Estimates conducted on Twitter report that, on average, social bots account for 9% to 15% of total active platform users [38]. This notable percentage is highly due to the crucial issue that bots *evolve* over time: in a large-scale experiment, it has been proved that neither Twitter admins, nor tech-savvy social media users, nor cutting-edge applications were able to tell apart evolving bots and legitimate users [8].

Academicians make their best efforts to fight the never ending plague of malicious bots populating social networks. The literature offers a plethora of successful approaches, based, e.g., on profile- [1, 7], network- [24, 39, 40], and posting-characteristics [5, 9, 17] of the accounts. In particular, the supervised approach proposed in [7] tested a series of rules and features from both the grey literature and the official publications on a reference dataset of genuine and fake accounts, leading to the implementation of a classifier which significantly reduces the cost for data gathering.

Remarkably, all the previous analyses rarely tackle the effect of random noise, which is indeed of utmost importance when studying complex systems. In [21], Jaynes showed how Statistical Physics could be derived from Information Theory from an entropy maximization principle. Following Jaynes work, in recent years the same approach has been extended to complex networks [6, 16, 26, 32, 33], to provide an unbiased benchmark for the analysis, by filtering out random noise. Such entropy-based null-models have demonstrated their effectiveness in reconstructing a network from partial information [31], in detecting early signals of structural changes [28, 34] and in assessing the systemic risk of a financial system [12, 19]. The approach is general and unbiased, being based on the concept of Shannon entropy. In a nutshell, starting from the real network, the method relies on three steps: 1. the definition of an *ensemble* of graphs; 2. the definition of the entropy for this ensemble and its maximization up to some (local or global) constraints [26]; 3. the maximization of the likelihood of the real network [16, 32].

In the present study, we merge the application of bot detection techniques with the use of an entropy-based null-model for the analysis of the content exchange on Twitter in the Italian debate about regulating the migration flux from Northern Africa. The corpus we analyzed resulted to be extremely informative in highlighting some otherwise hidden features of the dissemination of information in that debate, as we are going to describe in the following.

2 RESULTS AND DISCUSSION

In order to get the political affiliation of users, we focused on the bipartite network in which the two layers represent verified and unverified users, respectively, and the (undirected) links label the interactions between the two classes. The main idea is to infer the inclination of users towards a political point of view from (a proxy of) their contacts: if two users share a great amount of followers and followees, they probably have a similar political polarization. Following the strategy of [28], we use the above mentioned entropy-based framework to project the bipartite network on the layer of verified users, whose account information is reliable. Verified users have been clustered into 3 main groups, see Fig. 1: one group includes government representatives, the right wing and the Movimento 5 Stelle party; a second group includes the Italian Democratic party; a third one includes NGOs, online and offline media, journalists and some VIPs. Confirming results presented in other studies [11, 25, 27, 29], the polarization of unverified users is particularly strong: they interact quite exclusively with accounts of a single community. Differently than in other studies [3], the interaction of unverified users with verified ones is limited, and affects only one half of the total amount of unverified users. This is probably due to the fact that we focus on a debate that is wider than an election campaign and that could stimulate exchanges between users who do not usually participate in political discussions. Thus, we iteratively assign group memberships to unverified users, based on the political affiliation of the majority of all their followers and followees. This procedure reduces the number of unpolarized accounts of more than 35%. Curiously, the ratio of bot accounts that remain unpolarized after the 'political contagion' is higher than the analogous for all users. In any case, in the following, we will see that users, automated or not, taking *effectively* part to the debate are mostly polarized.

Finally, we extract the non trivial content exchange by adopting the validated projection developed in [3]: this permits to detect the significant flow of messages among users, discounting, at the same time, the virality of messages, the retweeting activity of users and their productivity in writing tweets. Such an approach provides the 'backbone' of the content exchange among users on Twitter.

The network represented in Fig. 2 is extremely informative for different reasons. The validated network contains only 14,883 validated users out of the 127,275 users in the dataset. This highlights the fact that just a minority of all users effectively contributes to the online debate on the migration flow. Interestingly, we found that the incidence of bots on the validated network is almost one third of the analogous measure on the entire dataset, signaling that the number of bots whose retweets are non compatible with a random activity is just a small minority. Since the target of a social bot is to increase audience of the online content of a specific user, such a reduction shows that the number of bots affecting significantly the political debate is limited.

The set of validated users is much more polarized than the whole set of users: we have that the overall fraction of unpolarized accounts represents more than 40% of all the accounts and more than 50% of the automated ones, while when considering the validated network, the same ratio is around 10% for the former and around 5% for the latter. Otherwise stated, the polarized bots pass the validation process more easily than their unpolarized counterparts and their contribution in spreading messages is more significant.

All the accounts that are mostly effective in delivering their messages (i.e., the Hubs [22]) refer to the blue area in Fig. 1, where we can find representatives of the the government in charge and the right wing. The first account referring to a community different from the blue one is the official account of the newscast 'TgLa7', at position 176th in the hub ranking.

The contribution of bots to the visibility of the various accounts shows that the fraction of bots that significantly retweet the content of two right wing political leaders (Mr. Salvini and Ms. Meloni) is greater than the incidence of bots in the validated network. Interestingly enough, other hubs show a smaller presence of bots among their followers, even if their hub score is not that different from the two political leaders.

Finally, we have that some hubs do share their bots: indeed we found non trivial overlap among the bots following the strongest hubs in the validated network; as mentioned before, the strongest hubs are from the right wing political area. To the best of our knowledge, this is the first time that such a behaviour is reported: in analyses tackling the same problem [35, 36, 38], only star-like sub-graphs were observed, with a big number of bots among the followers of a (presumably) human user. We underline that the considered shared bots are particularly effective, since they are validated by the entropy-based projection. Actually, the group of "right wing" bots, supporting at the same time various human accounts, is not the only one in the set, but it is the greatest: if we consider the subgraphs of human accounts sharing their bots, the former has 172 nodes against 58 of the latter. Moreover the first subgraph is by far more efficient; indeed, in the second one the greatest hub score ranks 176th.

It is well known that bots aim at increasing popularity of users by retweetting their messages, see, e.g., [8]: exactly what is revealed by the entropy-based filtering. The latter turns out to be extremely helpful, since *it hits one feature of an automated account that cannot be avoided by programmers.* To the best of our knowledge, the study

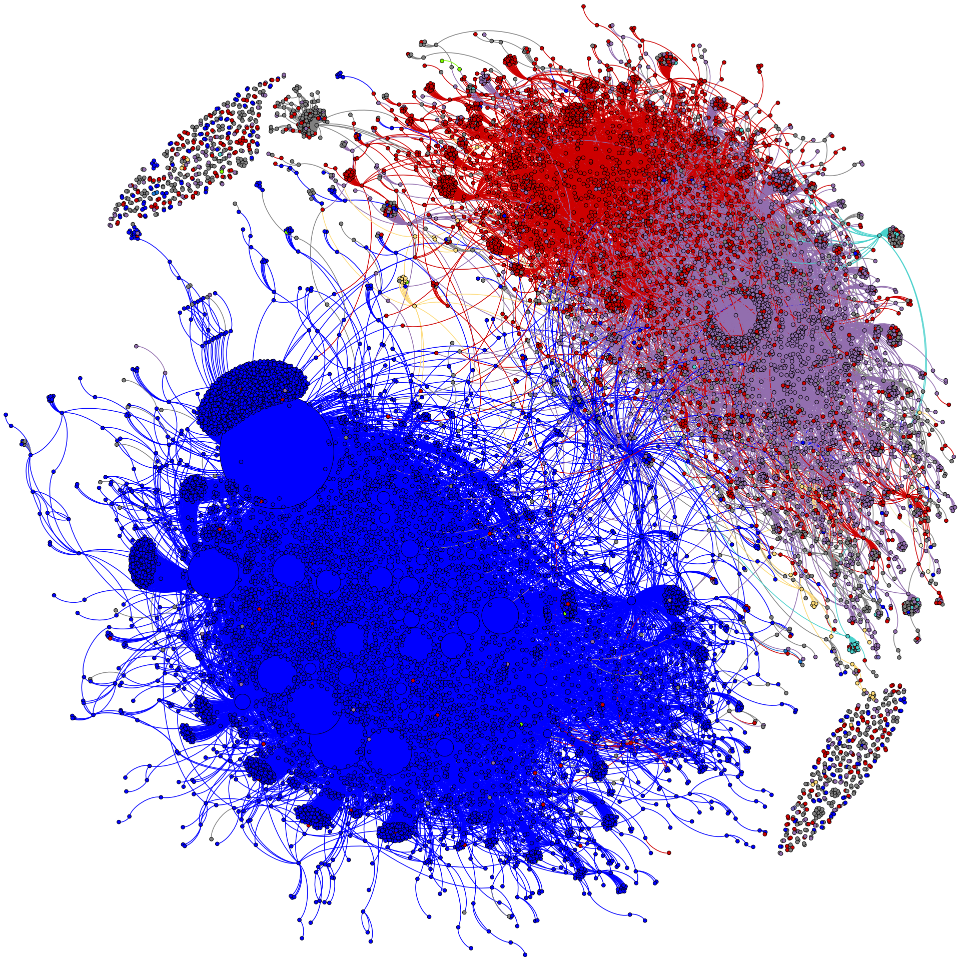

Figure 2: The validated network of content exchange; nodes have been colored according to the community of the previous 1. The dimension of the nodes is proportional to their hub score.

here presented is the first investigation that merges bot detection and entropy-based analysis of Twitter traffic. Moreover, the obtained results are in line with the previous work of [30], where the authors showed how bots massively support the spread of (low credibility) content. At the same time, the present outcome contributes in a different way, being not specifically focused on fake news, whereas [30] concentrates on the way fake news become viral. Interestingly enough, among the many studies of the 2016 US presidential election, Grinberg *et al.* [18] analyzed the proliferation of fake news on Twitter and determined both fake news spreaders and exposed users. Remarkably, it was found that fake news was 'most concentrated among conservative voters'. The role of bots in effectively conveying a message - for the first time here highlighted even in a 'shared fashion' - and the spreading of fake news in online debates of great importance [18, 30] leads us to a promising future direction of study, which include a deeper analysis of the exchanged messages, like the extraction of their sentiment and the contained mentions.

REFERENCES

[1] Prudhvi Ratna Badri Satya, Kyumin Lee, Dongwon Lee, Thanh Tran, and Jason Jiasheng Zhang. 2016. Uncovering fake likers in online social networks. In *CIKM*. ACM.

[2] Marco T Bastos and Dan Mercea. 2017. The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review* (2017), 0894439317734157.

[3] Carolina Becatti, Guido Caldarelli, Renaud Lambiotte, and Fabio Saracco. 2019. Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections. (jan 2019). arXiv:1901.07933 https://arxiv.org/abs/1901.07933http://arxiv.org/abs/1901.07933

[4] Alexandre Bovet and Hernán A. Makse. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nat. Commun.* 10, 1 (2019). https://doi.org/10.1038/s41467-018-07761-2 arXiv:1803.08491

[5] Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. 2016. DeBot: Twitter Bot Detection via Warped Correlation.. In *ICDM*. 817–822.

[6] Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and Guido Caldarelli. 2018. The Statistical Physics of Real-World Networks. *Nat. Rev. Phys.* 1, 1 (jan 2018), 58–71. https://doi.org/10.1038/s42254-018-0002-6 arXiv:1810.05095

[7] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2015. Fame for sale: efficient detection of fake Twitter followers. *Decision Support Systems* 80 (2015), 56–71.

[8] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2017. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In *Proceedings of the 26th International Conference on*

World Wide Web Companion (WWW'17). ACM, 963–972.

[9] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2018. Social fingerprinting: detection of spambot groups through DNA-inspired behavioral modeling. *IEEE Transactions on Dependable and Secure Computing* 15, 4 (2018), 561–576.

[10] Stefano Cresci, Fabrizio Lillo, Daniele Regoli, Serena Tardelli, and Maurizio Tesconi. 2019. Cashtag Piggybacking: Uncovering Spam and Bot Activity in Stock Microblogs on Twitter. *TWEB* 13, 2 (2019), 11:1–11:27. https://dl.acm.org/citation.cfm?id=3313184

[11] Michela Del Vicario, Fabiana Zollo, Guido Caldarelli, Antonio Scala, and Walter Quattrociocchi. 2017. Mapping social dynamics on Facebook: The Brexit debate. *Soc. Networks* 50 (2017), 6–16. https://doi.org/10.1016/j.socnet.2017.02.002 arXiv:arXiv:1610.06809v1

[12] Domenico Di Gangi, Fabrizio Lillo, and Davide Pirino. 2018. Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. *J. Econ. Dyn. Control* 94 (sep 2018), 117–141. https://doi.org/10.1016/j.jedc.2018.07.001 arXiv:1509.00607

[13] Emilio Ferrara. 2015. Manipulation and abuse on social media. *ACM SIGWEB Newsletter* Spring (2015), 4.

[14] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The Rise of Social Bots. *Commun. ACM* 59, 7 (June 2016), 96–104. https://doi.org/10.1145/2818717

[15] C. Gangware and W. Nemr. 2019. Weapons of Mass Distraction: Foreign State-Sponsored Disinformation in the Digital Age.

[16] Diego Garlaschelli and Maria I Loffredo. 2008. Maximum likelihood: Extracting unbiased information from complex networks. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 78, 1 (2008), 1–5. https://doi.org/10.1103/PhysRevE.78.015101 arXiv:cond-mat/0609015

[17] Maria Giatsoglou, Despoina Chatzakou, Neil Shah, Alex Beutel, Christos Faloutsos, and Athena Vakali. 2015. ND-Sync: Detecting Synchronized Fraud Activities. In *PAKDD*.

[18] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Political science: Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363, 6425 (jan 2019), 374–378. https://doi.org/10.1126/science.aau2706

[19] Stanislao Gualdi, Giulio Cimini, Kevin Primicerio, Riccardo Di Clemente, and Damien Challet. 2016. Statistically validated network of portfolio overlaps and systemic risk. *Sci. Rep.* 6 (mar 2016), 39467. https://doi.org/10.1038/srep39467 arXiv:1603.05914

[20] Mengdie Hu, Shixia Liu, Furu Wei, Yingcai Wu, John Stasko, and Kwan-Liu Ma. 2012. Breaking News on Twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2751–2754. https://doi.org/10.1145/2207676.2208672

[21] E.T. Jaynes. 1957. Information Theory and Statistical Mechanics. , 181–218 pages. https://doi.org/10.1103/PhysRev.106.620 arXiv:arXiv:1011.1669v3

[22] Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *J. ACM* (1999). https://doi.org/10.1145/324133.324140

[23] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a Social Network or a News Media?. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. ACM, New York, NY, USA, 591–600. https://doi.org/10.1145/1772690.1772751

[24] Shenghua Liu, Bryan Hooi, and Christos Faloutsos. 2017. HoloScope: Topology-and-Spike Aware Fraud Detection. In *CIKM*. ACM.

[25] Dimitar Nikolov, Diego F. M. Oliveira, Alessandro Flammini, and Filippo Menczer. 2015. Measuring Online Social Bubbles. (2015). https://doi.org/10.7717/peerj-cs.38 arXiv:1502.07162

[26] Juyong Park and Mark E J Newman. 2004. Statistical mechanics of networks. *Phys. Rev. E* 70, 6 (dec 2004), 66117. https://doi.org/10.1103/PhysRevE.70.066117

[27] Walter Quattrociocchi, Guido Caldarelli, and Antonio Scala. 2014. Opinion dynamics on interacting networks: Media competition and social influence. *Sci. Rep.* 4 (2014). https://doi.org/10.1038/srep04938

[28] Fabio Saracco, Riccardo Di Clemente, Andrea Gabrielli, and Tiziano Squartini. 2016. Detecting early signs of the 2007âĂŞ2008 crisis in the world trade. *Sci. Rep.* 6 (jul 2016), 30286. https://doi.org/10.1038/srep30286 arXiv:1508.03533

[29] Ana Lucía Schmidt, Fabiana Zollo, Antonio Scala, Cornelia Betsch, and Walter Quattrociocchi. 2018. Polarization of the vaccination debate on Facebook. *Vaccine* 36, 25 (2018), 3606–3612. https://doi.org/10.1016/j.vaccine.2018.05.040

[30] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature Communications* 9, 1 (2018), 4787. https://doi.org/10.1038/s41467-018-06930-7

[31] Tiziano Squartini, Guido Caldarelli, Giulio Cimini, Andrea Gabrielli, and Diego Garlaschelli. 2018. Reconstruction methods for networks: The case of economic and financial systems. https://doi.org/10.1016/j.physrep.2018.06.008 arXiv:1806.06941

[32] Tiziano Squartini and Diego Garlaschelli. 2011. Analytical maximum-likelihood method to detect patterns in real networks. *New J. Phys.* 13 (2011). https://doi.org/10.1088/1367-2630/13/8/083001 arXiv:arXiv:1103.0701v2

[33] Tiziano Squartini and Diego Garlaschelli. 2017. Maximum-entropy networks. Pattern detection, network reconstruction and graph combinatorics. *Springer* (2017), 116.

[34] Tiziano Squartini, Iman van Lelyveld, and Diego Garlaschelli. 2013. Early-warning signals of topological collapse in interbank networks. *Sci. Rep.* 3 (2013), 3357. https://doi.org/10.1038/srep03357 arXiv:10.1038/srep03357

[35] Massimo Stella, Marco Cristoforetti, and Manlio De Domenico. 2018. Influence of augmented humans in online interactions during voting events. (mar 2018). arXiv:1803.08086 http://arxiv.org/abs/1803.08086

[36] Massimo Stella, Emilio Ferrara, and Manlio De Domenico. 2018. Bots sustain and inflate striking opposition in online social systems. (2018). arXiv:1802.07292 http://arxiv.org/abs/1802.07292

[37] TNS opinion & social and Directorate-General Communications. 2018. Media use in the European Union. https://doi.org/10.2775/116707

[38] Onur Varol, Emilio Ferrara, Clayton A Davis, Filippo Menczer, and Alessandro Flammini. 2017. *Online Human-Bot Interactions: Detection, Estimation, and Characterization.* Technical Report. https://doi.org/10.1016/S0009-9260(05)80025-X arXiv:arXiv:1703.03107v2

[39] Binghui Wang, Neil Zhenqiang Gong, and Hao Fu. 2017. GANG: Detecting Fraudulent Users in Online Social Networks via Guilt-by-Association on Directed Graphs. In *ICDM*. IEEE.

[40] Shuhan Yuan, Xintao Wu, Jun Li, and Aidong Lu. 2017. Spectrum-based deep neural networks for fraud detection. In *CIKM*. ACM.