Fact-checking strategies to limit urban legends spreading in a segregated society

Marcella Tambuscio[1,2]* and Giancarlo Ruffo[1]

*Correspondence:
marcella.tambuscio@gmail.com
[1]Computer Science Department,
University of Turin, Corso Svizzera
185, 10149 Torino, TO, Italy
[2]Present address: Austrian Center
for Digital Humanities, Austrian
Academy of Sciences,
Sonnenfelsgasse 19, 1010 Wien,
Austria

Abstract

We propose a framework to study the spreading of urban legends, i.e., false stories that become *persistent* in a local popular culture, where social groups are naturally segregated by virtue of many (both mutable and immutable) attributes. The goal of this work is identifying and testing new *strategies* to restrain the dissemination of false information, focusing on the role of network polarization. Following the traditional approach in the study of information diffusion, we consider an epidemic network-based model where the agents can be 'infected' after being exposed to the urban legend or to its debunking depending on the belief of their neighborhood. Simulating the spreading process on several networks showing different kind of segregation, we perform a *what-if analysis* to compare strategies and to understand where it is better to locate *eternal fact-checkers*, nodes that maintain their position as debunkers of the given urban legend. Our results suggest that very few of these strategies have a chance to succeed. This apparently negative outcomes turns out to be somehow surprising taking into account that we ran our simulations under a highly pessimistic assumption, such that the 'believers', i.e., agents that accepted as true the urban legend after they have been exposed to it, will not change their belief no matter of how much external or internal additional informational sources they access to. This has implications on policies that are supposed to decide which strategy to apply to stop misinformation from spreading in real world networks.

Keywords: Misinformation spreading, Network segregation, Debunking strategies, Epidemics

Introduction

Our goal is to investigate new strategies to limit false news spreading, specially in presence of existing structural, geographical and/or social barriers. If segregation seems to be an intrinsic feature of modern urban environments, which policies can be implemented to empower fact-checking platforms? At the best of our knowledge there are some researches about identifying influential spreaders of information (Kitsak et al. 2010; Ghosh and Lerman 2010) or rumors (Borge-Holthoefer et al. 2013), but to our knowledge few efforts have been devoted to assess and compare possible debunking strategies exploiting the network topology. In this paper, we suggest the application of a what-if analysis based on epidemic modeling in order to explore new fact-checking policies to limit the diffusion of urban legends. This methodology is particularly useful in contexts where we do not have data on how a given strategy to restrain misinformation would perform,

because such action plans have never been applied in real life (or whose results have not been disclosed to scholars yet).

Online information consumption and polarization

Recently misinformation has been largely discussed (Lazer et al. 2018; Vosoughi et al. 2018) because it can imply serious consequences to our lives: even if in some cases fake news are intentionally disseminated to manipulate public opinion, there is a large amount of persistent rumors, or *urban legends*, that look as simple popular stories but are often related with social problems and leverage on fears, prejudices and emotions of people (Campion-Vincent 2017; Heath et al. 2001). In this framework, digital technologies as online social networks can facilitate the spreading of misinformation, specially because they are *homophily-driven*, built with the intent to connect like-minded people and often exhibit the presence of *echo chambers*, highly segregated environments with low content diversity and high degree of repetition (Adamic and Glance 2005; Conover et al. 2011; Pariser 2011; Bozdag and van den Hoven 2015). Moreover, these platforms involve filtering algorithms (DeVito 2017) and recommendation systems that give disproportionate visibility to popular content within social circles. These mechanisms of *algorithmic personalization* have been largely debated in literature to understand if they affect the evolution of opinions (Rossi et al. 2018; Bressan et al. 2016) and polarize the network (Perra and Rocha 2019; Dandekar et al. 2013; Geschke et al. 2019), or if, conversely, they do not have a leading role in the formation of echo chambers (Möller et al. 2018; Bakshy et al. 2015).

Segregation, homophily, and network topologies

Empirical analysis confirmed that online conversations involving misinformation appear to be highly polarized (Del Vicario et al. 2016; Bessi et al. 2015), but research about the role of the underlying topology of the network in information diffusion suggests that the level of segregation can affect the spreading in different ways (Tambuscio et al. 2018; Bakshy et al. 2012; Onnela et al. 2007; Weng et al. 2013; Nematzadeh et al. 2014). However, many attributes or factors that lead to the formation of segregated communities are somehow 'mutable': for example, nodes that join or leave the network can contribute to create new shortest paths to otherwise distant communities, interests change during time and, as a consequence, attention to given topics. On the other hand, segregation has been largely studied (Oka and Wong 2015; Massey and Denton 1993; 1987) and observed (Bajardi et al. 2015; Herdağdelen et al. 2016; Lamanna et al. 2018) in urban environments, involving features of human life as language, religion, ethnicity, education, employment and so on. Many of these attributes are 'immutable', and the topology of the network can be shaped accordingly. The theoretical framework provided by the Schelling model (Schelling 2006) shows that spatial segregation is somehow natural even in tolerant societies: in a simple grid where agents can change their place if the percentage of similar individuals in their spatial proximity is lower than a given percentage, even a small bias towards homophily, but still highly tolerant w.r.t. diversities, leads to totally segregated configurations. Interestingly, these patterns have been observed in real societies (Gracia-Lázaro et al. 2009; Clark and Fossett 2008). In our research, to better generalize our findings, we focus on different network topologies that can be caused by social dynamics such as preferential attachment, as well as intrinsic

segregation patterns that are dependent on immutable characteristics of the population of a city.

Information spreading modeling

The tradition of information (and consequently, rumor and misinformation) diffusion modeling has involved different approaches: epidemic models (Moreno et al. 2004; Daley and Kendall 1964), influence models (Goldenberg et al. 2001; Granovetter and Soong 1983), opinion dynamics (Castellano et al. 2009) are the most known. In particular researchers distinguish among biological *simple contagion* (induced by a single exposure, as epidemic models) and *complex contagion* (dependent on multiple exposures, as influence models) (Centola and Macy 2007). Even if complex contagion has been found to well describe observed information cascades and predict their size (Lerman 2016; Mønsted et al. 2017; Romero et al. 2011), the complexity of the phenomenon seems to involve other factors (Min and San Miguel 2018; Zhuang et al. 2017) and many models based on epidemics have been proposed to study rumor and misinformation (Zhao et al. 2013; Jin et al. 2013; de Arruda et al. 2016). Moreover, in models based on complex contagion agents have only one possibility to activate their neighbors and never de-activate, meaning that they do not take into account forgetting mechanisms. Considering misinformation spreading, this is an important element to represent, since many psychological studies suggest it has a significant role (Lewandowsky et al. 2012; Nyhan and Reifler 2010).

Our contribution

Following the epidemic approach we extended a compartmental model (Tambuscio et al. 2015) where the agents can be in one of the three states: Susceptible, if they ignore the news, Believer, if they support the urban legend, or Fact Checker, if they decide to foster the debunking. Evolution in time is given by transition rates that allow an agent to change state, and these rates depend on the following parameters: the number of neighbors Believers or Fact-Checker, the spreading rate β (common to both hoax and debunking), the credibility of the hoax α (that gives some priority to misinformation but also can represent different propensities to believe), the forgetting rate p_f (the probability for agents in both Believer and Fact-Checker states to return to the Susceptible state). Since it is known that bias and personal beliefs often prevent people to look for fact-checking (Nyhan and Reifler 2010; Lewandowsky et al. 2012), we consider here the worst case in the framework given by the model in Tambuscio et al. (2015), when no one verify the information (meaning that it is not possible to switch from Believer to Fact-Checker state) and the debunking spreads only as a competing opinion of the rumor.

We simulate spreading dynamics on three types of networks: *scale free networks*, networks formed by *communities* characterized by different values of credibility (including a simulation on the well known *'polblogs' real network*) and grid configurations obtained by means of the *Schelling segregation model*. The parameter α in the model represents the credibility of the hoax, i.e. the tendency that each agent has to believe to it. This is an advantage for the misinformation piece, reflecting the results of several psychological studies (Allport and Postman 1947; DiFonzo and Bordia 2007; Silverman 2015) that indicate credibility (combined with repetition) is a strong enhancer for rumor diffusion; the effect is even amplified if the story matches pre-existing beliefs (*confirmation bias*) (Nyhan and Reifler 2010). Then it is reasonable to represent urban legends having some

priority respect to their fact-check, at least for some communities. Under these conditions, in absence of a transition Believer - Fact Checker (no verifying activity), the rumor at some point affects the whole population and the debunking dies out, even for very low values of α. Therefore, to limit the propagation of the rumor in such a configuration we propose here to fix some nodes to be *eternal* Fact Checkers (i.e., that never return to be Susceptible) and then we run several simulations to compare a group of strategies targeting different type of nodes as *eternal fact checkers*. For instance, if the network is highly segregated, a solution to be tested would be to place fact-checkers on the frontier among the clusters, so that we could exploit natural segregation to isolate the urban legend only in some clusters. Nevertheless, if the frontier is not totally covered, the rumor can eventually go beyond it and propagate in the whole network. In this case, if the same number of fact-checkers is placed in the higher degree nodes (hubs), the rumor diffusion can be partially limited. We will discuss in detail these strategies through simulations of the model in different network topologies, highlighting the fact that in each case we are able to find a strategy that contains the misinformation: these findings can be useful in proposing new policies to foster debunking and fight fake news spreading.
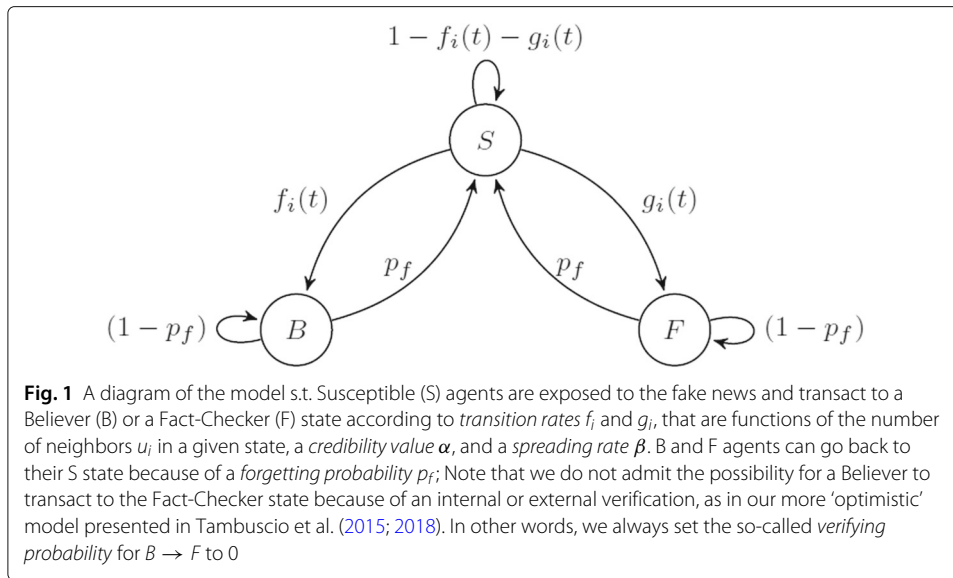
The model

Agents can be in three states: Susceptible (ignoring the news), Believer (believe it is true) or Fact Checker (believe it is false), and the state can change according to transition rates. In particular:

- assuming that a Susceptible agent can decide to believe in either the hoax or the fact checking as a result of interaction over interpersonal ties (Rosnow and Fine 1976), the rumor/debunking spreading (transitions $S \rightarrow B$ and $S \rightarrow F$) depends on the number of believers/fact-checkers among neighbors and on a parameter α that represents the *credibility* of the legend;
- Believers and Fact Checkers can return to the Susceptible state with a fixed forgetting probability p_f (transitions $B \rightarrow S$ and $F \rightarrow S$);

Please observe that this model is a 'pessimistic' variation of a previous model (Tambuscio et al. 2015), that also follows the traditional approach of epidemic spreading (Moreno et al. 2004) to understand misinformation diffusion dynamics; in fact, in the previous model, we introduced also the possibility for an agent to switch from Believer to Fact Checker with a given verifying probability p_v (Tambuscio et al. 2015), meaning that the debunking can also be spread by external factors (online fact-checking platforms, for instance). On the other hand we consider here the *worst possible scenario* to test our strategies: users do not (want to) verify the news, they can only be influenced by their Fact-Checker neighbors (if any) when they are in their Susceptible state. After they get a position, they can only return to the Susceptible state if they forget what they learnt about the news they have been previously exposed to.

Let us describe formally the model (also shown in Fig. 1). At each time t, each agent can be in one of the three states: Susceptible (S), Believer (B) or Fact Checker (F), denoted by a state indicator $s_i(t) = \left(s_i^S, s_i^B, s_i^F\right)$. For example, if the node 2 at time t is in the Believer state, we will have $s_2(t) = (0, 1, 0)$. The triplet $p_i(t) = \left[p_i^S, p_i^B, p_i^F\right]$ is the vector of probabilities that node i is in any of the states at time t. The dynamics of the system at time $t+1$ will be then given by a random realization of p_i at $t + 1$:

Fig. 1 A diagram of the model s.t. Susceptible (S) agents are exposed to the fake news and transact to a Believer (B) or a Fact-Checker (F) state according to *transition rates* f_i and g_i, that are functions of the number of neighbors u_i in a given state, a *credibility value* α, and a *spreading rate* β. B and F agents can go back to their S state because of a *forgetting probability* p_f; Note that we do not admit the possibility for a Believer to transact to the Fact-Checker state because of an internal or external verification, as in our more 'optimistic' model presented in Tambuscio et al. (2015; 2018). In other words, we always set the so-called *verifying probability* for $B \rightarrow F$ to 0

$$p_i^B(t+1) = f_i(t)s_i^S(t) + (1-p_f)s_i^B(t) \tag{1}$$

$$p_i^F(t+1) = g_i(t)s_i^S(t) + (1-p_f)s_i^F(t) \tag{2}$$

$$p_i^S(t+1) = p_f s_i^B(t) + p_f s_i^F(t) + \left[1 - f_i(t) - g_i(t)\right]s_i^S(t) \tag{3}$$

where p_f is the *forgetting* probability, whereas f_i and g_i are the transition rates from S to B and F respectively. These functions depend on the number of neighbors that are Believers and Fact Checkers at time t, denoted, for each agent i, by $n_i^B(t)$ and $n_i^F(t)$. The functions f_i and g_i are defined as follows:

$$f_i(t) = \beta \, \frac{n_i^B(t)(1+\alpha)}{n_i^B(t)(1+\alpha) + n_i^F(t)(1-\alpha)} \tag{4}$$

$$g_i(t) = \beta \, \frac{n_i^F(t)(1-\alpha)}{n_i^B(t)(1+\alpha) + n_i^F(t)(1-\alpha)} \tag{5}$$

where $\beta \in [0,1]$ is the spreading rate and $\alpha \in [0,1)$ represents the credibility of the legend (meaning that is more believable when α is close to 1) and give some priority to the piece of misinformation respect to the debunking. Please observe that when $\alpha = 0$ the hoax is still able to spread, but there is not any advantage over the fact-checking.

Once the model has reached the equilibrium, we denote with S_∞, B_∞ and F_∞ the asymptotic density of agents in the three states.

In other words, we are representing an urban legend spreading with an opinion dynamics model where the hoax compete with its debunking at a local level of the social interaction of agents. Please notice that this model follows a SIS-like (Susceptible-Infected-Susceptible) dynamics where the Infected state is split into Believer and Fact Checker compartments.

The networks

The goal of this work is comparing differnt strategies to limit the misinformation spreading in a segregated society; in our simulations we will consider different types of networks that exhibit several degrees of segregation. Let us briefly recap what we know about the

role of the topology network in the model described in "The model" section. The original version of the model, where the agents can verify a piece of information and switch from Believer to Fact Checker, showed the same behavior on random, scale-free and real networks (Tambuscio et al. 2015). The work described in Tambuscio et al. (2018) focused on the evolution of the model dynamics in networks formed by two communities, one made of more *gullible* agents, while the other is set to be more *skeptic*: these communities exhibit different propensities to believe (different values of α): extensive simulations show that the segregation level of the network can both help to spread or stop the misinformation, depending on the forgetting rate. In particular these networks were artificially generated rewiring two random networks, obtaining different levels of segregation.
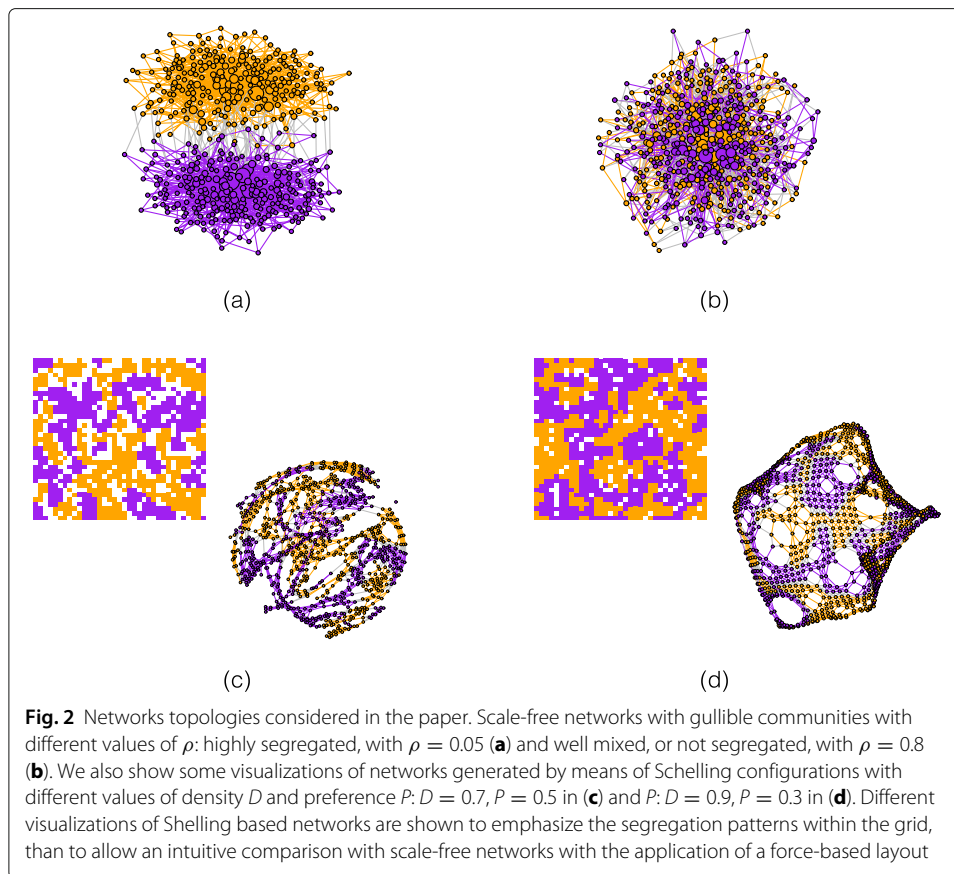
In the following lines we will introduce the networks on which we tested our experiments and debunking strategies to perform our what-if analysis with our new model.

Synthetic scale-free networks

First, we run our model on scale free networks artificially generated by means of a Barabasi-Albert process with $m = 3$. The networks we created have $N = 1000$ nodes and $M \approx 3000$ links, with a mean degree $\langle k \rangle \approx 6$. Second, we consider, as in Tambuscio et al. (2018), a network formed by two clusters where the agents have different tendencies to believe (α), but we start with two totally separated *scale free* networks and then we rewire them randomly. More specifically, at the beginning ($\rho = 0$) we have two completely segregated networks that represent the two communities (the gullible one has a higher tendency to believe, $\alpha = 0.8$, whereas the skeptic has a lower one, $\alpha = 0.3$). We consider a parameter $\rho \in [0, 1]$ and perform $\rho \cdot M$ rewiring trials keeping the degree distribution of the nodes, where M is the total number of edges of the networks. When $\rho \approx 1$ the obtained network is well mixed, meaning that half of the links connect the two communities. In this framework, we will explore what happens fixing some nodes as eternal Fact-Checkers ($p_f = 0$), choosing them at random, among the nodes with highest degree (hubs), or among the nodes on the boundary between the two communities. To allow a comparison with the first case we started with two networks of 500 nodes, then the resulting networks have the same number of nodes ($N = 1000$) and links ($M \approx 3000$) of the ones considered before. In Fig. 2a-b there are two examples of networks generated with different values of ρ.

Observations: In our previous works Tambuscio et al. (2015; 2018) we ran our simulations with varying values of N. We understood that when N is larger (up to 10,000), the general behavior does not change, hence we kept N smaller to run many different realizations of the model faster.

It is also important to observe that scale-free artificial networks showing different segregation values could have been generated by means of configuration models. In our comparative what-if analysis we used three different topologies (BA graphs, Shelling based networks, and PolBlog as a network built up from real data), and we observed comparable behaviors, that led us to conclude it was not necessary to simulate our strategies on another family of artificially generated graphs. Nevertheless, it is true that configuration models have less drawbacks in terms of non-trivial correlations than BA networks, and so an additional analysis can be performed as a future work.

Fig. 2 Networks topologies considered in the paper. Scale-free networks with gullible communities with different values of ρ: highly segregated, with $\rho = 0.05$ (**a**) and well mixed, or not segregated, with $\rho = 0.8$ (**b**). We also show some visualizations of networks generated by means of Schelling configurations with different values of density D and preference P: $D = 0.7$, $P = 0.5$ in (**c**) and P: $D = 0.9$, $P = 0.3$ in (**d**). Different visualizations of Shelling based networks are shown to emphasize the segregation patterns within the grid, than to allow an intuitive comparison with scale-free networks with the application of a force-based layout

Real networks

We considered a real network (POLBLOGS) between weblogs on US politics, collected during US elections in 2004 (Adamic and Glance 2005). We chose this network because is formed by two labeled communities that reflect somehow an opinion (belief) of the nodes, and we considered them as gullible and skeptic groups described before, assigning them different values of credibility. Specifically, we used a modified version of the original network: we mapped the directed graph to an undirected one, we selected the largest connected component, lowering the number of vertices from 1490 to 1222, and finally we removed all multi-edges and loops, lowering the number of edges from 16725 to 16714.

Synthetic Schelling networks

Finally, we considered the Schelling segregation model as a simple representation of a segregated urban environment (Schelling 2006). In this model two groups of agents are randomly placed on a grid of size S. The number of agents N is obtained by a density D as $N = D * S^2$. A parameter P denotes the *preference*, i.e., the desired fraction of neighbors of the same type for all the agents. These preference can also be seen as an inverse measure of tolerance (a lower preference corresponds to higher number of neighbors of a different type). Clearly in a random configuration there will be some unsatisfied agents: at each time t they move to an empty cell. Running the simulations, different configurations can be obtained: when there is an equilibrium (all the agents are satisfied) the network results to be segregated in small communities of the same group Fig. 2c-d.

This class of topologies is interesting because the spatial segregation arises from the local effect of homophily based on more 'immutable' characteristics of the individual (i.e., ethnicity, religion, language, and so on): urban networks can be shaped very differently w.r.t. a (on-line) social network. Indeed, even for a low value of P we can obtain these segregated configurations, see Fig. 2c-d. Here, we run the Schelling model with $S = 35$ so that, varying D in $[0.7, 0.9]$ we have $N \approx 1000$. We consider the final configuration of the model as the starting one for our simulations: the two groups represent the *gullible* and *skeptic* agents, and they are assigned with different values of α.
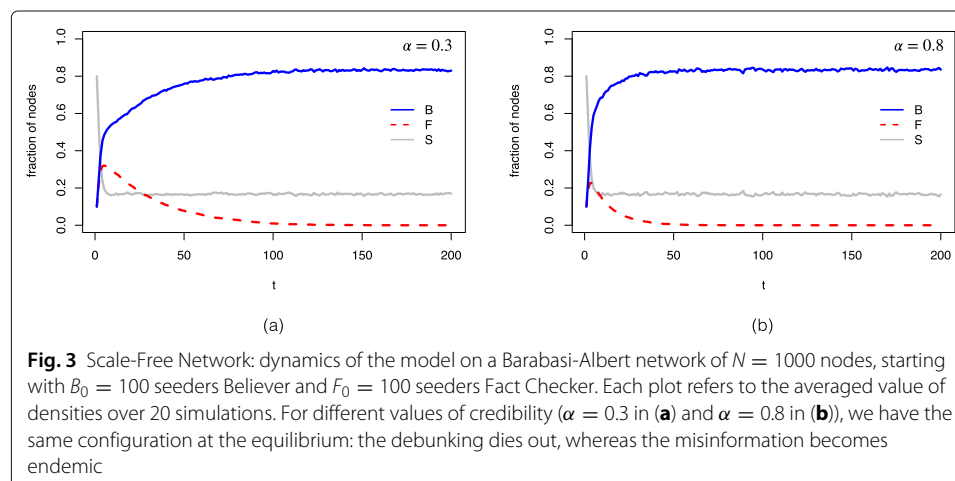
Results

In our model agents can become Believer or Fact-Checker only if they are infected by their neighbors, then we start our simulations with a population of Susceptible agents and some Believer and Fact-Checker seeders ($B_0 = F_0 = 0.1 * N$). To understand the behavior of the model in different configurations we performed extensive numerical simulations fixing for simplicity the spreading rate $beta = 0.5$ and the forgetting probability $p_f = 0.1$.

Scale-free networks without gullible communities

First of all we consider a simple Barabasi-Albert network with N=1000 nodes and we set the Believer and Fact Checker seeders at random. In this configuration, the urban legend wins over its debunking that completely disappears ($F_\infty = 0$), even for low values of credibility (see Fig. 3). Neither increasing the number of Fact Checker seeders helps: the curve of debunkers goes to 0 in any case, but more slowly. Notice that α does not affect the asymptotic number of Believers, B_∞. This is because, in our model, the asymptotic number of Susceptible agents depends only on p_f and β (it can be easily obtained by the Eq. 3, for details see Tambuscio et al. (2015)): then, if $F_\infty = 0$, also B_∞ does not depend on the credibility, that affects only the rate of Believers and Fact-Checkers in the global spread.

Then, our goal is to explore strategies to limit, even if not defeating, the hoax spreading. The main idea is to fix some nodes as *eternal* Fact Checkers (eFC), meaning that some of them never forget their belief ($p_f = 0$). On the other hand, it is reasonable to think that there are some individuals that do not believe to the hoax and that will never change their opinion. We tested two ways of choosing these individuals: randomly and
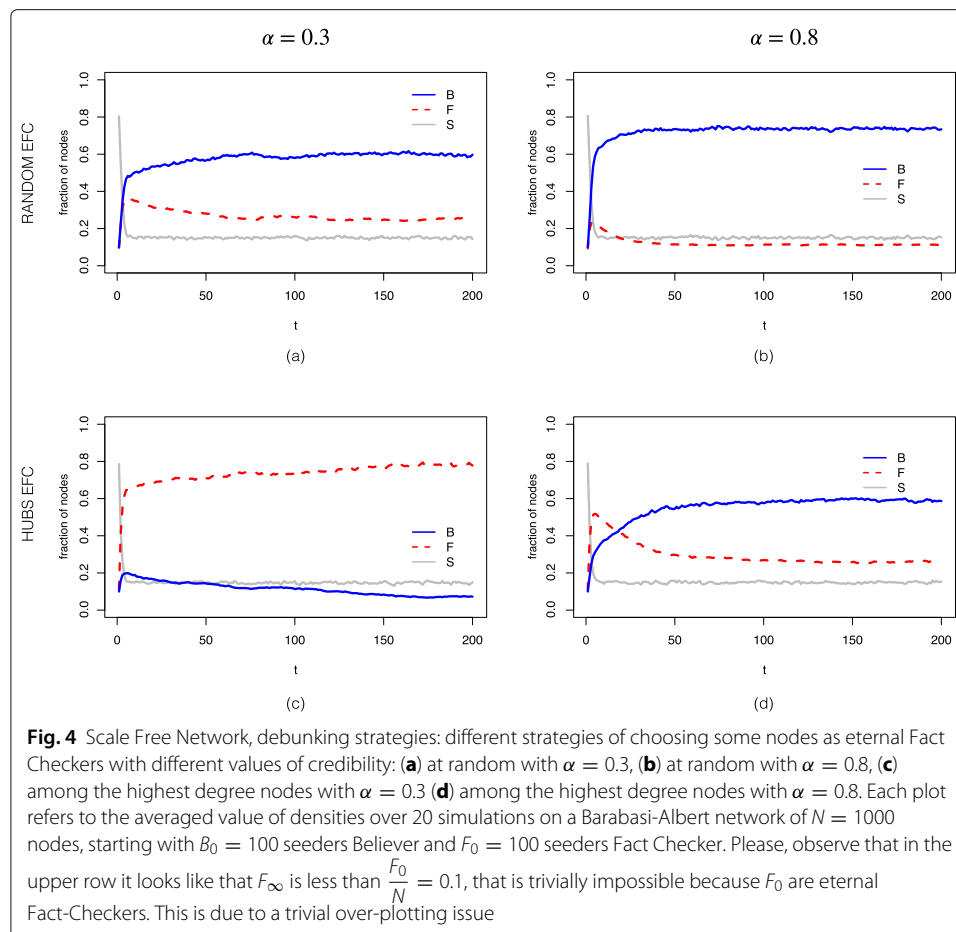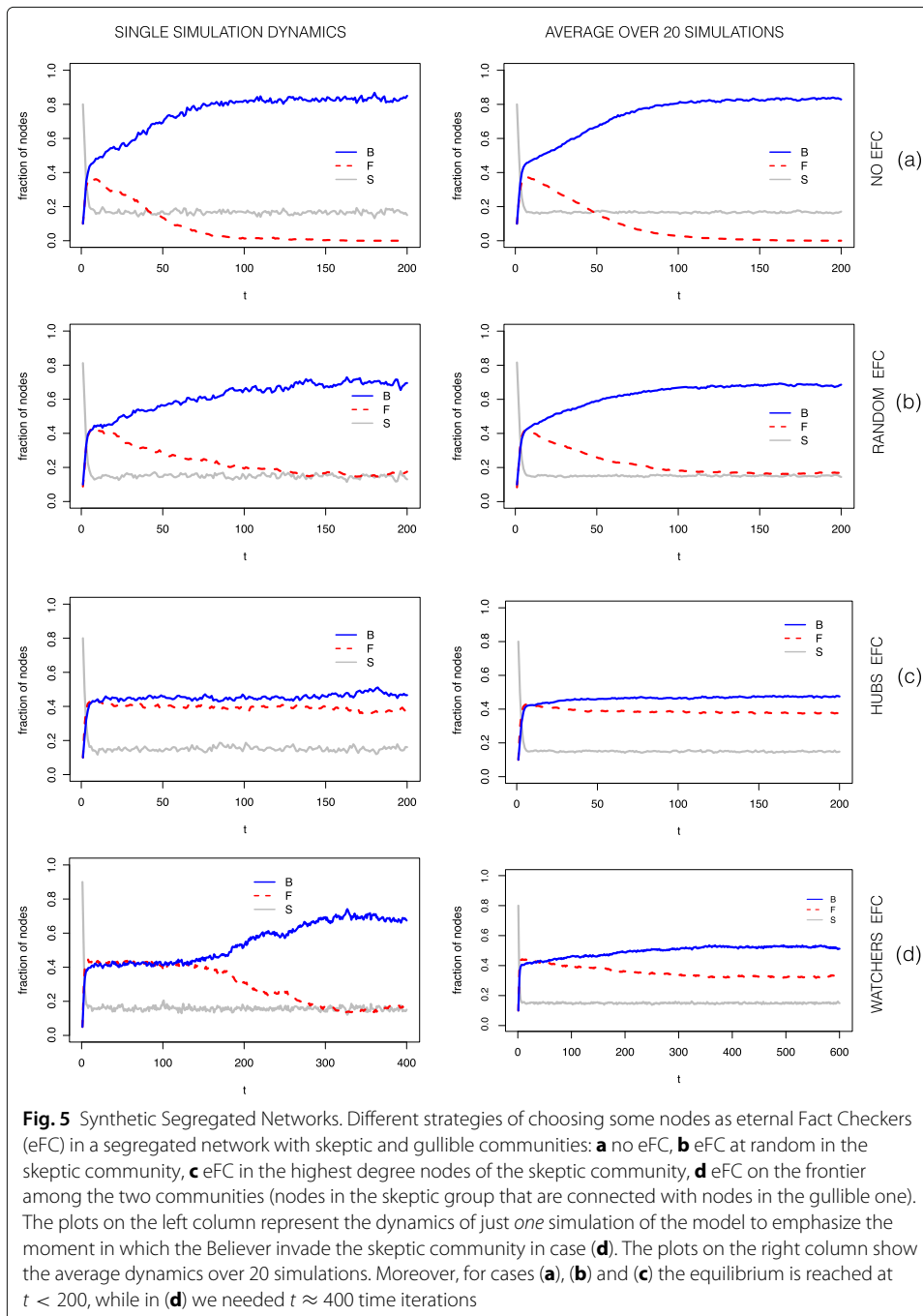

Fig. 3 Scale-Free Network: dynamics of the model on a Barabasi-Albert network of $N = 1000$ nodes, starting with $B_0 = 100$ seeders Believer and $F_0 = 100$ seeders Fact Checker. Each plot refers to the averaged value of densities over 20 simulations. For different values of credibility ($\alpha = 0.3$ in (**a**) and $\alpha = 0.8$ in (**b**)), we have the same configuration at the equilibrium: the debunking dies out, whereas the misinformation becomes endemic

among the highest degree nodes (hubs). For simplicity we set the number of eternal Fact Checkers equal to F_0, meaning that they are exactly the seeders. In the first case we set randomly in the network the eternal Fact Checkers and we obtain a small but important effect ($F_\infty > 0$) in the case of low credibility (see Fig. 4a) that limits the hoax spreading. However, with higher values of credibility, the misinformation still reaches the whole network (see Fig. 4b) and the Fact Checkers die out (only the eFC survive at the equilibrium). In the second case we consider the first F_0 highest degree nodes: this strategy is absolutely winning. For low values of credibility the hoax is almost eradicated (see Fig. 4c), while for high values of credibility there is a substantial reduction and in both cases the debunking becomes endemic at equilibrium (see Fig. 4d). Clearly the strategy is promising, suggesting that a systematic activity of debunkers can be useful even if people do not verify the news, just because they create a spreading dynamics on their own.

Scale-free networks with gullible communities

We consider here the case in which there is a gullible community, as proposed in Tambuscio et al. (2018). The network is built as described in "The networks" section, rewiring two Barabasi-Albert networks of $N/2$ nodes that represent *gullible* and *skeptic* communities: to characterize them we set $\alpha = 0.8$ for the first group, meaning that the legend is highly credible for them, and $\alpha = 0.3$ for the second, meaning that the legend is

Fig. 4 Scale Free Network, debunking strategies: different strategies of choosing some nodes as eternal Fact Checkers with different values of credibility: (**a**) at random with $\alpha = 0.3$, (**b**) at random with $\alpha = 0.8$, (**c**) among the highest degree nodes with $\alpha = 0.3$ (**d**) among the highest degree nodes with $\alpha = 0.8$. Each plot refers to the averaged value of densities over 20 simulations on a Barabasi-Albert network of $N = 1000$ nodes, starting with $B_0 = 100$ seeders Believer and $F_0 = 100$ seeders Fact Checker. Please, observe that in the upper row it looks like that F_∞ is less than $\frac{F_0}{N} = 0.1$, that is trivially impossible because F_0 are eternal Fact-Checkers. This is due to a trivial over-plotting issue

not so credible for the agents in this group. At $t = 0$ there are $B_0 = 0.1 * N$ Believer seeders in the gullible community and $F = 0.1 * N$ Fact Checker seeders in the skeptic one. The results are reported in Fig. 5: we included both the single simulation and the averaged dynamics to emphasize the moment when, during a single realization placing the Fact Checkers on the frontier (Fig. 5d), the Believers 'invade' the skeptic group. This 'invasion' is something that eventually happens, even if in some single realizations this event appears after many time iterations: indeed it is interesting to notice that the equilibrium in this case is reached after significantly more time iterations respect to other cases.

Fig. 5 Synthetic Segregated Networks. Different strategies of choosing some nodes as eternal Fact Checkers (eFC) in a segregated network with skeptic and gullible communities: **a** no eFC, **b** eFC at random in the skeptic community, **c** eFC in the highest degree nodes of the skeptic community, **d** eFC on the frontier among the two communities (nodes in the skeptic group that are connected with nodes in the gullible one). The plots on the left column represent the dynamics of just *one* simulation of the model to emphasize the moment in which the Believer invade the skeptic community in case (**d**). The plots on the right column show the average dynamics over 20 simulations. Moreover, for cases (**a**), (**b**) and (**c**) the equilibrium is reached at $t < 200$, while in (**d**) we needed $t \approx 400$ time iterations
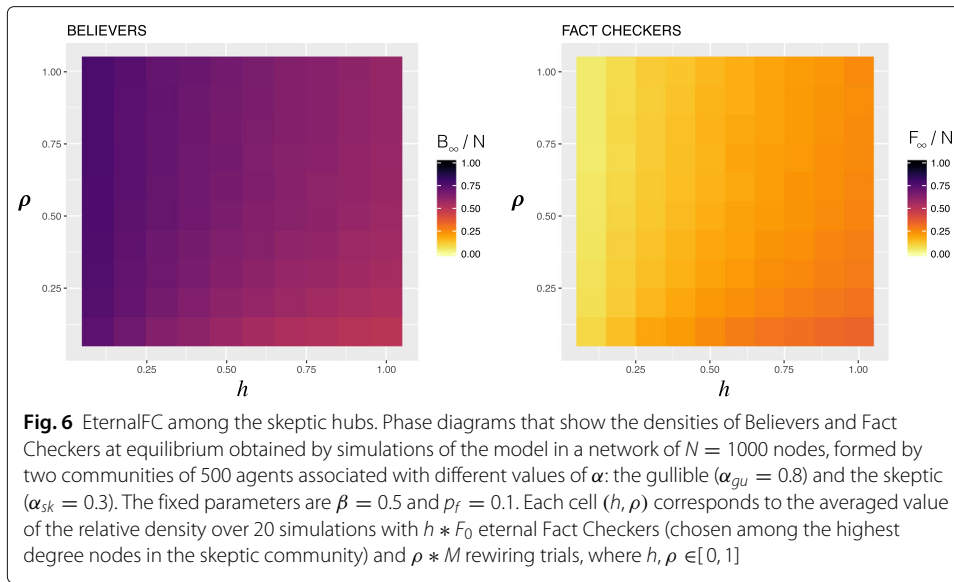
Running the simulation without debunking strategies (see Fig. 5a) the misinformation can conquer all the network: even a highly segregated community (half of the entire population) formed by skeptic people with low tendency to believe to an hoax is not enough to limit the urban legend spreading. Then we fix, as in the previous case, some skeptic nodes as eternal Fact Checker and we try three different strategies to choose them: randomly, among the highest degree nodes, among the nodes on the frontier. As before, in all the three cases, for simplicity we set the number of eternal Fact Checkers equal to F_0: in the case of the frontier if the network is highly segregated and we saturated all the possible nodes we choose the remaining eternal Fact Checkers at random in the skeptic community.

In the first case, setting the eternal Fact Checkers at random (see Fig. 5b), we can see that the misinformation is partially limited, but it still reaches the skeptic community and stays endemic at the equilibrium. The second strategy, choosing the hubs (see Fig. 5c), has indeed an important effect, limiting the global spreading of the hoax (that is now basically confined in the gullible community) and it guarantees a constant high number of Fact Checker in the skeptic community (with a slightly better result for the real network, that can depend on the segregation level, as we will see next). Finally, the third case, locating the eternal debunkers on the frontier among the communities, is very interesting, because two events can occur. If the frontier is totally covered by the eternal Fact Checker, trivially we have the best possible result in this framework: the misinformation is totally confined in the gullible community and the skeptic one is totally protected by its "watchers" (look at the first time iterations of first column plot in Fig. 5d). But, even if there is only one possible "door", at some point the Believer agents can invade the skeptic community and at the equilibrium we have number of endemic Fact Checker very similar to the first case (see first column plot in Fig. 5d when $t \approx 150$). We would like to remark that here we are considering a toy-network in a borderline case of two communities more or less connected: the nodes on the frontier then represent the *bridges* of our network, and, indeed, they exhibit high values of betweenness centrality on average, specially for higher segregation.

Therefore the most powerful strategy would be the third, but only if it is possible to cover all the frontier with the eternal Fact Checkers. However, since we are exploring possible solutions to limit misinformation in the real world where new links form continuously and keeping a community totally protected is not achievable, we can conclude that the best strategy, among the proposed ones, is the second, fixing eternal Fact Checkers among the hubs of the network.

However, we simulated only the case of $F_0 = 0.1 * N$ eternal Fact Checkers: we want to explore now what happens if we consider just a fraction of F_0 hubs ($h \in [0, 1]$) or nodes on the frontier ($w \in [0, 1]$). Moreover, we are also interested in how the segregation affects the number of Believer and Fact Checker agents at equilibrium, then we also consider a range of values for $\rho \in [0, 1]$, the parameter that rules the number of rewiring trials (see "The networks" section). Then we run several simulations showing the results in Figs. 6 and 7. In the first figure each tile $[h, \rho]$ represents the density at equilibrium of Believers $\left(\frac{B_\infty}{N} \right)$ or Fact Checkers $\left(\frac{F_\infty}{N} \right)$ on the average of 20 simulations with $h * F_0$ eternal Fact Checker agents and $\rho * M$ rewiring trials. We remind that the network is more segregated when $\rho \approx 0$. Similarly, in the second for $[w, \rho]$. In both cases (Figs. 6 and 7) we can observe that increasing the number of skeptic hubs/watchers fixed as eternal Fact Checkers helps a little bit in limiting the hoax spreading. Specifically, comparing the two

Fig. 6 EternalFC among the skeptic hubs. Phase diagrams that show the densities of Believers and Fact Checkers at equilibrium obtained by simulations of the model in a network of $N = 1000$ nodes, formed by two communities of 500 agents associated with different values of α: the gullible ($\alpha_{gu} = 0.8$) and the skeptic ($\alpha_{sk} = 0.3$). The fixed parameters are $\beta = 0.5$ and $p_f = 0.1$. Each cell (h, ρ) corresponds to the averaged value of the relative density over 20 simulations with $h * F_0$ eternal Fact Checkers (chosen among the highest degree nodes in the skeptic community) and $\rho * M$ rewiring trials, where $h, \rho \in [0, 1]$

figures, it is clear that the hubs strategy is the most effective one, since the number of Fact-Checkers at equilibrium is significantly higher. On the other hand, we notice that increasing the number of links among the communities (i.e. increasing ρ) causes a larger spreading of the misinformation in the skeptic community (and consequently a smaller number of Fact Checkers). This means (once again) that the role of network segregation is absolutely nontrivial: in the version of the model with verifying probability $\neq 0$ (Tambuscio et al. 2018) it was found that segregation indeed helps the spread of the hoax for low forgetting rates. Here we obtain, apparently surprisingly, the opposite result: network segregation helps the debunking. But this is not a conflicting results: if the gullible agents have the possibility to fact-check, the debunking will eventually spread also in the gullible community and, when this group is more isolated, the hoax

Fig. 7 EternalFC among the skeptic watchers. Phase diagrams that show the densities of Believers and Fact Checkers at equilibrium obtained by simulations of the model in a network of $N = 1000$ nodes, formed by two communities of 500 agents associated with different values of α: the gullible ($\alpha_{gu} = 0.8$) and the skeptic ($\alpha_{sk} = 0.3$). The fixed parameters are $\beta = 0.5$ and $p_f = 0.1$. Each cell (w, ρ) corresponds to the averaged value of the relative density over 20 simulations with $w * F_0$ eternal Fact Checkers (chosen among the skeptic nodes that are connected with gullible ones) and $\rho * M$ rewiring trials

has more probabilities to survive. Conversely, when agents do not have the possibility to verify, the debunking can only be spread from the skeptic group, then if the two communities are more polarized, it has more chances to not be eradicated from the network.
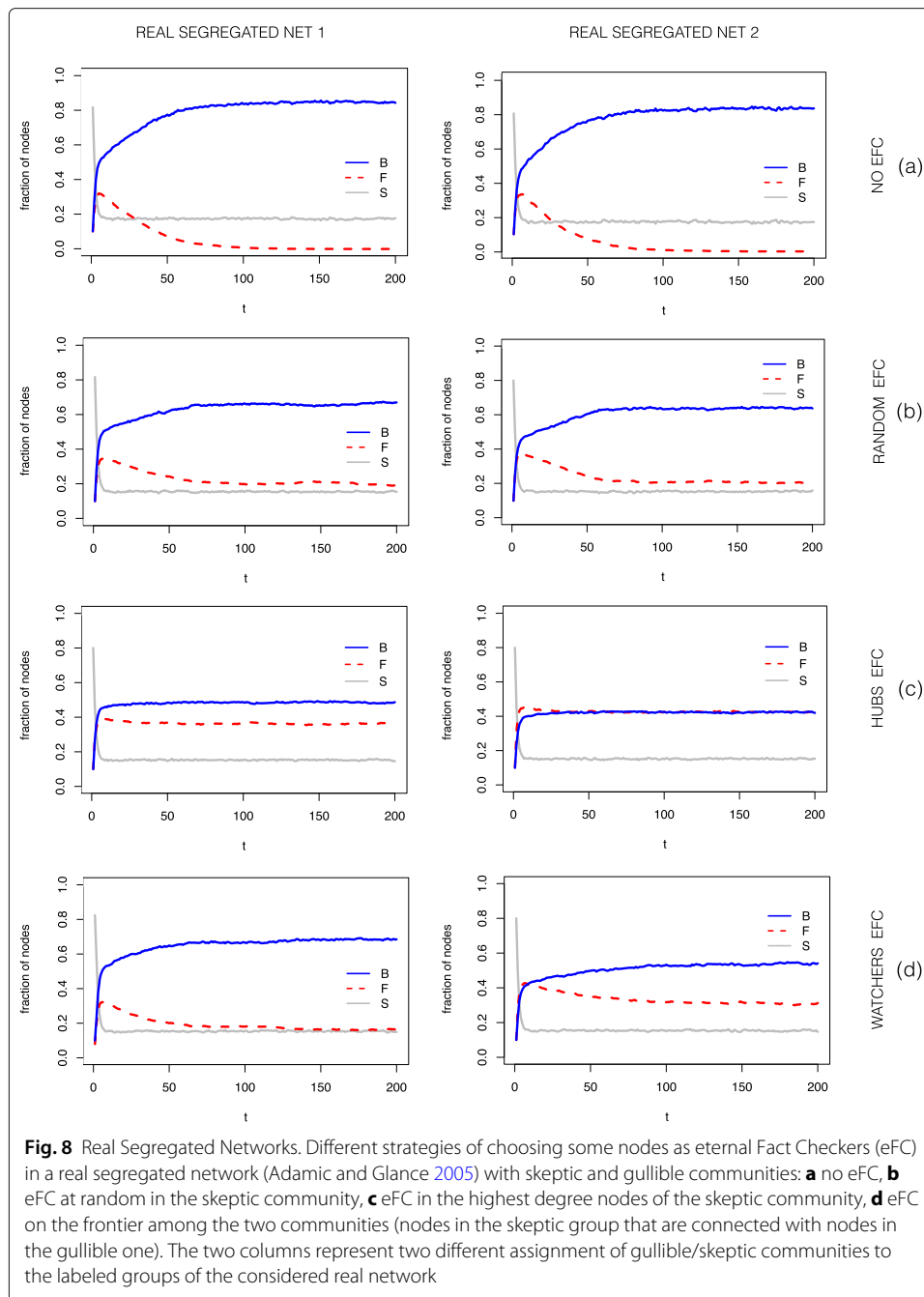
Real-world segregated networks

To test our model not only on synthetic networks, we also considered a real network with similar features, the network POLBLOGS described in Sec. 2. This network exhibits two highly segregated communities, and the nodes are labeled with their political alignment (conservative or liberal). We produced two configurations to test our model, considering first the two labeled groups as the gullible and the skeptic communities and then vice-versa: Fig. 8 show the results of these simulations. Even if not completely identical, both cases reflect the same behavior and the hubs strategy is again the best one possible: in the second configuration we observe better outcomes for Fact-Checkers for all the strategies, and this is probably due to the fact that the two communities does not have the same size, as in the simulations (one is formed by 636 nodes, the other one by 586). Moreover, comparing these plots with Fig. 5, it is easy to see that simulations on synthetic and real segregated networks produce analogous results, specially in the second case, for which we have an extraordinary similarity.

Schelling model networks

In this section we consider another type of segregated network: the grid configuration of the Schelling segregation model after the equilibrium is reached (see Sec. 2). This agent based model showed that segregation is something that can arise even in very tolerant contexts and has been used for instance to study residential segregation of ethnic groups: empirical evidences supporting Schelling-like patterns were observed between the Jewish and Arabs communities in Israel (Hatna and Benenson 2012). The Schelling grids at equilibrium provide us a framework to test the hoax spreading model and its debunking strategies on a segregated urban environment where the topology of the network is inherently shaped by social and human attributes that historically led to separate and isolate groups (ethnicity, religion, gender, language etc).
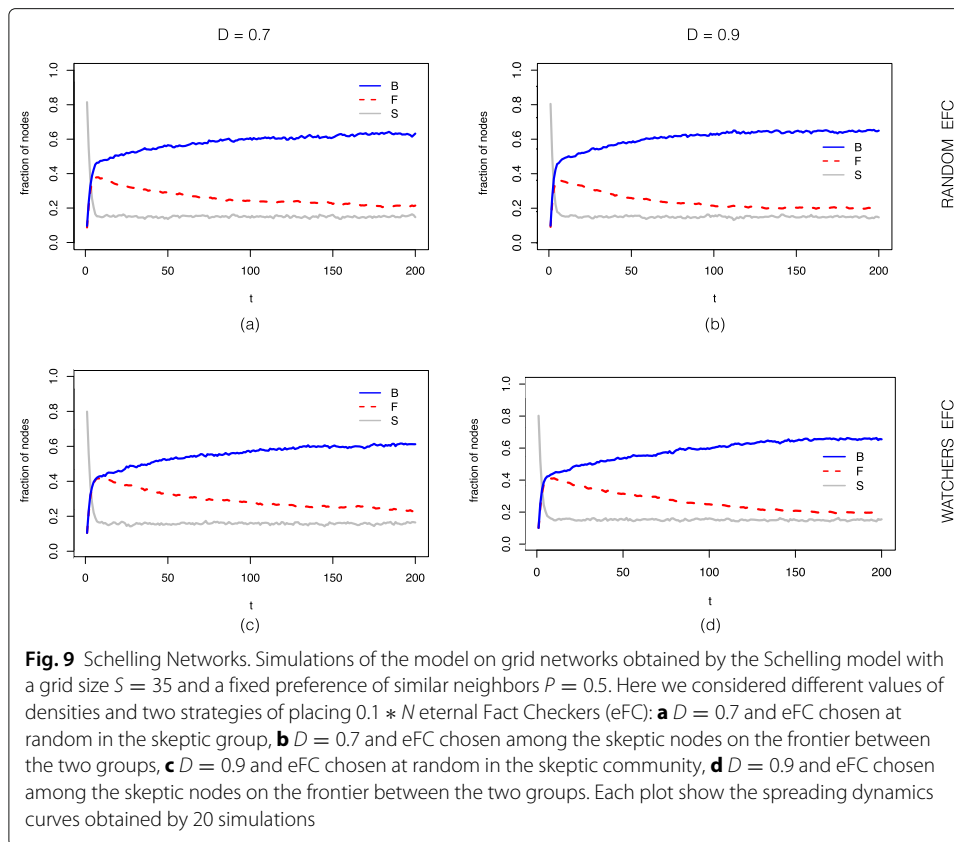
We assigned different values of α to the agents of the two groups to determine gullible/skeptic communities and we placed the Believer and Fact Checker seeders among the gullible and the skeptic respectively. Figure 9 shows the dynamics evolution for different values of densities D and preferences P. We chose values of preference P and density D that guarantee a largest connected component formed by more than 95% of the nodes N, so that we can exclude trivial cases in which there are totally isolated communities in which of course the hoax or the debunking survive without competitors. As in the previous cases, the hoax easily wins over the debunking in the spreading competition, reaching all the network, then we try to fix some nodes as eternal Fact Checkers. Here we do not consider the strategy of placing these nodes in the hubs, simply because it is meaningless on a grid. Then, let us consider the possibility to locate the eternal debunkers randomly in the skeptic network or to place them on the frontiers. Surprisingly, both strategies can limit partially the misinformation spreading, but with much less evident effect respect to the scale-free networks. We set the Schelling configuration with $S = 35$ so that, varying D in $[0.7, 0.9]$ we have $N \approx 1000$, and we considered an

Fig. 8 Real Segregated Networks. Different strategies of choosing some nodes as eternal Fact Checkers (eFC) in a real segregated network (Adamic and Glance 2005) with skeptic and gullible communities: **a** no eFC, **b** eFC at random in the skeptic community, **c** eFC in the highest degree nodes of the skeptic community, **d** eFC on the frontier among the two communities (nodes in the skeptic group that are connected with nodes in the gullible one). The two columns represent two different assignment of gullible/skeptic communities to the labeled groups of the considered real network

average value $P = 0.5$ for the preference of similar neighbors. In Fig. 9 it is quite clear that the debunking becomes endemic (in the skeptic community) but it does not disappear. The strategies have the same effect: the legend spreading is restrained a little bit and the debunking becomes endemic (in the skeptic community) but it does not disappear. The second strategy reaches the equilibrium more slowly (see Fig. 9b and d).

Discussion

Summarizing, we focused on the worst scenario provided by a misinformation spreading model, based on epidemics, in which agents can be infected by an urban legend or by its

Fig. 9 Schelling Networks. Simulations of the model on grid networks obtained by the Schelling model with a grid size $S = 35$ and a fixed preference of similar neighbors $P = 0.5$. Here we considered different values of densities and two strategies of placing $0.1 * N$ eternal Fact Checkers (eFC): **a** $D = 0.7$ and eFC chosen at random in the skeptic group, **b** $D = 0.7$ and eFC chosen among the skeptic nodes on the frontier between the two groups, **c** $D = 0.9$ and eFC chosen at random in the skeptic community, **d** $D = 0.9$ and eFC chosen among the skeptic nodes on the frontier between the two groups. Each plot show the spreading dynamics curves obtained by 20 simulations

debunking, then they can forget their belief about it and turn to be infected again; nevertheless, our pessimistic assumption is that once agents opted to become Believers, they will not verify the news anymore keeping their belief (until they forget their position). We defined this as the worst possible scenario because the fact-checking can only be broadcast trough neighbors contagion, meaning that the debunking platforms and activities could appear useless and inefficient. Indeed, not surprisingly, under these assumptions our simulations show that the hoax easily becomes endemic and the debunking disappears. In order to limit misinformation, this is a quite negative results: this is reflected by some other relevant studies that show how fact checking can be not effective and sometimes counterproductive (Butler et al. 2011; Nyhan and Reifler 2010; Lewandowsky et al. 2012), while the hoaxes proliferate creating highly polarized communities in the communication networks (Del Vicario et al. 2016; Bessi et al. 2015).

Nevertheless, keeping this pessimistic scenario, we tested some fact checking strategies that involve the introduction of *eternal fact-checkers*, agents that support the debunking and never forget their belief: the location of these agents has a crucial role in shaping the global diffusion of the urban legend and its debunking at equilibrium.

First, our simulation results on scale-free networks show that fixing the highest degree nodes as the eternal fact-checkers is the more successful strategy in limiting the hoax spreading, while choosing them randomly has a lower effect (even if the debunking at equilibrium survives, trivially in the immediate neighborhood of the eternal fact-checkers).

Second, in order to explore the role of segregation, we run our simulations on two types of polarized networks where the segregation is achieved in two different ways.

- In the first case, scale-free networks (synthetic and real) formed by two communities more or less segregated, the winning strategy is (again) to fix the skeptic hubs as eternal debunkers. This is more powerful than fixing them at random or on the frontier, that would be the most powerful strategy only when it is totally covered by eternal fact-checkers but this is clearly not affordable, since the real networks are dynamics. Indeed, in this case our simulations highlight that eventually the hoax finds a way to "dig through the wall" and spread in the other community, becoming endemic in it even if the agents in this group are more skeptic, i.e. they are less likely to believe to the urban legend. Then the frontier approach has the same outcome of the random one, both in synthetic and real segregated networks. Moreover, we find that in this framework of the model (when there is not verifying activity) the segregation of the network can restrain the misinformation spread because it prevents that the hoax spread in the skeptic community. For comparison with the framework in which agents can verify the news (then some Believer turn into Fact Checker) see Tambuscio et al. (2015; 2018).
- In the second case, the network is obtained by a realization of the Schelling model, i.e. it is a grid and every node has a low degree ($k \leq 8$) then we can not consider hubs. Nevertheless, fixing some eternal fact-checkers (at random or on the frontiers among the groups) works as well in limiting the legend spreading.

To draw a conclusion from our experimental settings, our what-if analysis show evidences that, even in a very pessimistic scenario where no one verify the news, some debunking strategies can be applied and have a partial success in limiting the misinformation spread, specially exploiting the presence of more skeptic agents in the network. Conversely, a censorship action on the nodes that broadcast hoaxes could not be helpful since new nodes can easily replace the silenced ones. Therefore, our results can surely be helpful in developing new policies to build fact-checking platforms and to foster their usage.

Conclusions

Misinformation is surely one of the most dangerous risks of our hyper-connected society and some proposed solutions involve the creation of accounts' blacklists or the development of tools to give less visibility to specific items labeled as fake news. Then interesting questions are arising (John Borthwick 2016): how to legislate without limiting freedom of speech and which authority should trust to an eventual law-making of the Internet? With this intent many fact-checking platforms have been proposed[1].

How these projects can become more effective? In this work we considered a simplified version of an epidemics-based model where the misinformation spreading is described only as a competition process among an urban legend and its debunking. The fact checking activity of the debunkers has been frequently labeled as useless and counterproductive because of psychological and social factors. Then we focused on the worst case scenario:

---

[1]http://www.disinfobservatory.org/, https://weverify.eu/, http://www.truly.media/

the agents can not verify the news, and the debunking can only be spread at a neighbor-hood level influencing agents that have not taken a position against or in favor a given fake news yet. We proved that, in different network topologies, the strategy of fixing the belief of a portion of the Fact-Checkers can indeed limit the misinformation spreading, even if the location of these agents has a big influence on the success of these strate-gies. This could mean that, even if the debunking services provided by the main stream media or online platforms are not much visited, they are still useful to restrain a fake news diffusion, specially if their usage is strategically coordinated by a skeptic community.

In the future we would like to collect data to better validate our model, developing a platform in which users can express their belief towards a news and some agents can be activated as eternal Fact-Checkers in strategic locations of the network. Moreover, on the theoretical side, we would like to explore next what happens on networks made of $n > 2$ communities with different propensities to believe.

We think that our findings, based on a what-if analysis that helped us to study a domain where we do not have enough data, can help to shed lights on the understanding of the complex phenomenon of misinformation spreading. Specifically, they can suggest new debunking policies to empower the existing fact-checking platforms or new social experiments in real contexts to test the proposed strategies and the role of segregation.

References
Adamic LA, Glance N (2005) The political blogosphere and the 2004 U.S, election: Divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery. ACM, Chicago. pp 36–43
Allport GW, Postman L (1947) The Psychology of Rumor. Henry Holt, Oxford, England
Bajardi P, Delfino M, Panisson A, Petri G, Tizzoni M (2015) Unveiling patterns of international communities in a global city using mobile phone data. EPJ Data Sci 4(1):3
Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news and opinion on facebook. Science 348(6239):1130–1132. https://doi.org/10.1126/science.aaa1160
Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: Proceedings of the 21st International Conference on World Wide Web. ACM, New York. pp 519–528

Bessi A, Petroni F, Del Vicario M, Zollo F, Anagnostopoulos A, Scala A, Caldarelli G, Quattrociocchi W (2015) Viral misinformation: The role of homophily and polarization. In: Proceedings of the 24th International Conference on World Wide Web. ACM, New York. pp 355–356

Borge-Holthoefer J, Meloni S, Gonçalves B, Moreno Y (2013) Emergence of influential spreaders in modified rumor models. J Stat Phys 151(1-2):383–393

Bozdag E, van den Hoven J (2015) Breaking the filter bubble: democracy and design. Ethics Inf Technol 17(4):249–265

Bressan M, Leucci S, Panconesi A, Raghavan P, Terolli E (2016) The limits of popularity-based recommendations, and the role of social ties. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. pp 745–754

Butler AC, Fazio LK, Marsh EJ (2011) The hypercorrection effect persists over a week, but high-confidence errors return. Psychon Bull Rev 18(6):1238–1244

Campion-Vincent V (2017) Rumor Mills: The Social Impact of Rumor and Legend. Routledge, Abingdon

Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. Rev Mod Phys 81(2):591

Centola D, Macy M (2007) Complex contagions and the weakness of long ties. Am J Sociol 113(3):702–734

Clark WA, Fossett M (2008) Understanding the social context of the schelling segregation model. Proc Natl Acad Sci 105(11):4109–4114

Conover M, Ratkiewicz J, Francisco M, Gonçalves B, Flammini A, Menczer F (2011) Political polarization on Twitter. In: Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM). AAAI, Barcelona

Daley DJ, Kendall DG (1964) Epidemics and rumours. Nature:1118. https://doi.org/10.1038/2041118a0

Dandekar P, Goel A, Lee DT (2013) Biased assimilation, homophily, and the dynamics of polarization. Proc Natl Acad Sci 110(15):5791–5796

de Arruda GF, Rodrigues FA, Rodriiguez PM, Cozzo E, Moreno Y (2016) Unifying markov chain approach for disease and rumor spreading in complex networks. arXiv preprint arXiv:1609.00682

Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W (2016) The spreading of misinformation online. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.1517441113

DeVito MA (2017) From editors to algorithms: A values-based approach to understanding story selection in the facebook news feed. Digit J 5(6):753–773

DiFonzo N, Bordia P (2007) Rumor Psychology: Social and Organizational Approaches. American Psychological Association, Washington

Geschke D, Lorenz J, Holtz P (2019) The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. Br J Soc Psychol 58(1):129–149

Ghosh R, Lerman K (2010) Predicting influential users in online social networks. In: SNA-KDD: PROCEEDINGS OF KDD WORKSHOP ON SOCIAL NETWORK ANALYSIS. ACM, New York

Goldenberg J, Libai B, Muller E (2001) Talk of the network: A complex systems look at the underlying process of word-of-mouth. Mark Lett 12(3):211–223

Gracia-Lázaro C, Lafuerza LF, Floría LM, Moreno Y (2009) Residential segregation and cultural dissemination: An axelrod-schelling model. Phys Rev E 80(4):046123

Granovetter M, Soong R (1983) Threshold models of diffusion and collective behavior. J Math Sociol 9(3):165–179

Hatna E, Benenson I (2012) The schelling model of ethnic residential dynamics: Beyond the integrated-segregated dichotomy of patterns. J Artif Soc Soc Simul 15(1):6

Heath C, Bell C, Sternberg E (2001) Emotional selection in memes: the case of urban legends. J Pers Soc Psychol 81(6):1028

Herdağdelen A, Adamic L, Mason W, et al. (2016) The social ties of immigrant communities in the united states. In: Proceedings of the 8th ACM Conference on Web Science. ACM, New York. pp 78–84

Jin F, Dougherty E, Saraf P, Cao Y, Ramakrishnan N (2013) Epidemiological modeling of news and rumors on twitter. In: Proceedings of the 7th Workshop on Social Network Mining and Analysis. ACM, New York. p 8

John Borthwick JJ (2016) A Call for Cooperation Against Fake News. https://medium.com/whither-news/a-call-for-cooperation-against-fake-news-d7d94bb6e0d4

Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. Nat Phys 6(11):888

Lamanna F, Lenormand M, Salas-Olmedo MH, Romanillos G, Gonçalves B, Ramasco JJ (2018) Immigrant community integration in world cities. PloS ONE 13(3):0191612

Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Metzger MJ, Nyhan B, Pennycook G, Rothschild D, et al. (2018) The science of fake news. Science 359(6380):1094–1096

Lerman K (2016) Information is not a virus, and other consequences of human cognitive limits. Future Internet 8(2):21

Lewandowsky S, Ecker U. K, Seifert C. M, Schwarz N, Cook J (2012) Misinformation and its correction continued influence and successful debiasing. Psychol Sci Public Interest 13(3):106–131

Massey DS, Denton NA (1987) Trends in the residential segregation of blacks, hispanics, and asians: 1970-1980. Am Sociol Rev:802–825

Massey DS, Denton NA (1993) American Apartheid: Segregation and the Making of the Underclass. Harvard University Press, Cambridge

Min B, San Miguel M (2018) Competing contagion processes: Complex contagion triggered by simple contagion. Sci Rep 8(1):10422

Möller J, Trilling D, Helberger N, van Es B (2018) Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. Inf Commun Soc 21(7):959–977

Mønsted B, Sapieżyński P, Ferrara E, Lehmann S (2017) Evidence of complex contagion of information in social media: An experiment using twitter bots. PloS ONE 12(9):0184148

Moreno Y, Nekovee M, Pacheco AF (2004) Dynamics of rumor spreading in complex networks. Phys Rev E 69(6):066130

Nematzadeh A, Ferrara E, Flammini A, Ahn Y-Y (2014) Optimal network modularity for information diffusion. Phys Rev Lett 113(8):088701

Nyhan B, Reifler J (2010) When corrections fail: The persistence of political misperceptions. Polit Behav 32(2):303–330

Oka M, Wong DW (2015) Spatializing segregation measures: An approach to better depict social relationships. Cityscape 17(1):97–114

Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci 104(18):7332–7336

Pariser E (2011) The Filter Bubble: What the Internet Is Hiding from You. Penguin, Penguin, UK

Perra N, Rocha LE (2019) Modelling opinion dynamics in the age of algorithmic personalisation. Sci Rep 9(1):7261

Romero DM, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the 20th International Conference on World Wide Web. ACM. pp 695–704

Rosnow RL, Fine GA (1976) Rumor and Gossip: The Social Psychology of Hearsay. Elsevier, New York

Rossi WS, Polderman JW, Frasca P (2018) The closed loop between opinion formation and personalised recommendations. arXiv preprint arXiv:1809.04644

Schelling TC (2006) Micromotives and Macrobehavior. WW Norton & Company, New York

Silverman C (2015) Lies, Damn Lies and Viral Content. Tow Center for Digital Journalism, Columbia University, New York. https://doi.org/10.7916/D8Q81RHH

Tambuscio M, Oliveira DF, Ciampaglia GL, Ruffo G (2018) Network segregation in a model of misinformation and fact-checking. J Comput Soc Sci 1(2):261–275

Tambuscio M, Ruffo G, Flammini A, Menczer F (2015) Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. In: Proceedings of the 24th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, New York. pp 977–982

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151

Weng L, Ratkiewicz J, Perra N, Gonçalves B, Castillo C, Bonchi F, Schifanella R, Menczer F, Flammini A (2013) The role of information diffusion in the evolution of social networks. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York. pp 356–364

Zhao L, Qiu X, Wang X, Wang J (2013) Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks. Phys A Stat Mech Appl 392(4):987–994

Zhuang Y, Arenas A, Yağan O (2017) Clustering determines the dynamics of complex contagions in multiplex networks. Phys Rev E 95(1):012312

Publisher's Note