

Controlling Contagion Processes in Activity Driven Networks

Suyu Liu,^{1,2} Nicola Perra,² Márton Karsai,² and Alessandro Vespignani^{2,3}

¹State Key Lab. of Industrial Control Technology, Institute of Cyber-systems and Control, Zhejiang University, Hangzhou 310027, China

²Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston MA 02115 USA

³Institute for Scientific Interchange Foundation, Turin 10133, Italy

(Dated: February 5, 2014)

The vast majority of strategies aimed at controlling contagion processes on networks consider the connectivity pattern of the system either quenched or annealed. However, in the real world, many networks are highly dynamical and evolve, in time, concurrently to the contagion process. Here, we derive an analytical framework for the study of control strategies specifically devised for a class of time-varying networks, namely activity-driven networks. We develop a block variable mean-field approach that allows the derivation of the equations describing the co-evolution of the contagion process and the network dynamic. We derive the critical immunization threshold and assess the effectiveness of three different control strategies. Finally, we validate the theoretical picture by simulating numerically the spreading process and control strategies in both synthetic networks and a large-scale, real-world, mobile telephone call dataset.

PACS numbers: 89.75.Hc, 64.60.aq

The spreading of infectious diseases, malwares, scientific ideas, and memes are just a few examples of real world phenomena that can be modeled as contagion processes on networks [1–4]. It has long been acknowledged that network structures and connectivity patterns are a relevant factor in determining the properties of spreading processes. Considering these findings, a number of strategies aimed at controlling contagion phenomena on the basis of the nodes' properties such as degree, k-core index, or betweenness have been proposed [5–7]. The efficiency of each strategy is measured by its effect on the contagion process when a fraction p of nodes is removed from the system. More precisely, the smaller the fraction p of nodes removed in order to halt the contagion process, the more effective the strategy.

Although most real world networks show a high level of dynamic activity, the large majority of theoretical results concerning the control of contagion processes have been obtained by using a complete timescale separation between the contagion process, τ_P , and the change in network's structure, τ_G . In these approaches, the dynamical process takes place in either static ($\tau_P \ll \tau_G$) or annealed ($\tau_P \gg \tau_G$) networks. However, when the two time scales are comparable, these convenient approximations might introduce uncontrolled biases in the characterization of the contagion process [8–32].

Here we investigate the effect of time-varying connectivity patterns on contagion control strategies by considering the specific class of activity driven network models [28]. In particular, we consider the susceptible-infected-susceptible (SIS) contagion model [33] and derive analytically its critical immunization threshold in three different control strategies. We also validate qualitatively the findings obtained in synthetic networks by studying the effect of each strategy in a large-scale mo-

bile telephone call dataset.

Activity driven models consider heterogenous populations where each node i is characterized by a specific activity rate a_i . This is the probability per unit time to establish interactions with other individuals. The activity rates are assigned according to a given probability distribution $F(a)$. The generative network process is defined according to the following rules (see Fig. 1-A-B): *i*) At each discrete time step t , the network G_t starts with N disconnected vertices; *ii*) With probability $a_i \Delta t$, each vertex i becomes active and generates m links that are connected to m other randomly selected vertices; *iii*) At the next time step $t + \Delta t$, all the edges in the network G_t are deleted. All interactions have a constant duration Δt . In the following, without loss of generality, we will set $\Delta t = 1$.

Activity driven networks are random and memoryless. The full dynamics of the network and its ensuing structure is completely encoded in the activity distribution $F(a)$. Moreover, integrating very large activity driven networks in finite time windows yield graphs characterized by degree distributions following the functional form $F(a)$, see the Supplementary Information (SI) and Ref. [28] for further details. We adopt heavy-tailed distributions of activity i.e. $F(a) = Ba^{-\gamma}$ with activities restricted in the region $a \in [\epsilon, 1]$ to avoid divergences for $a \rightarrow 0$, which approximate the behavior observed in real data for a number of real-world networks [28, 29] and more generally are found in human behavioral patterns. The following calculations can be easily extended to any functional form of the activity distribution, as shown in the SI. Although activity driven models in their simplest formulation do not account for many features such as link persistence, homophily, and underlying social structures [8–23], they allow the analytical formulation of the concurrent network and contagion process dy-

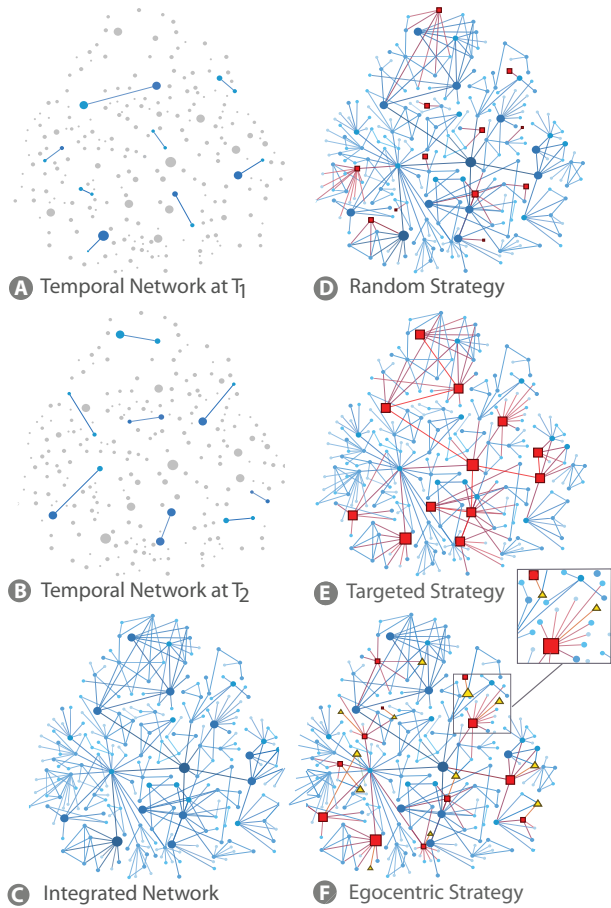


FIG. 1: Schematic representation of activity driven model and control strategies. Panels A and B show the temporal network at two different time steps T_1 and T_2 . Panel C presents the integrated network over a certain period of time. The size and color of each node describes its activity, while the width and color of each link describes the weight. Panels D, E, and F show random, targeted, and egocentric control strategy respectively. Immunized nodes are marked in red; probes in yellow.

namics in the form of appropriate mean-field equations, thus allowing the analytical study of the dynamical process of interest [25, 28–32].

In the following, we study contagion processes using the basic SIS model [33]. Each node at each time t can be in the susceptible, S^t , or infectious, I^t , state. The basic SIS rules thus define a reaction scheme of the type $S + I \rightarrow 2I$ with probability per unit time β and $I \rightarrow S$ with probability per unit time μ , which represent the contagion and recovery processes, respectively. We consider the case in which these two probabilities are the same for all the nodes, and defer the analysis of scenarios characterized by heterogeneous probability distributions for future works. A central concept of contagion phenomena is the epidemic threshold. It defines the conditions necessary for the spreading of the disease. In networks, the threshold depends on the moments of the

degree distribution $P(k)$ that specify the probability that any node is connected to k distinct nodes. In uncorrelated annealed networks, the threshold condition reads $\frac{\beta}{\mu} \geq \frac{\langle k \rangle^2}{\langle k^2 \rangle}$, where $\langle k \rangle$ and $\langle k^2 \rangle$ are the first and second moment of the degree distribution, respectively [11]. In this expression, $\beta = \lambda \langle k \rangle$ takes into account the average contacts per node $\langle k \rangle$ and the per contact probability of transmission λ . While the expression might be different in the case of static networks [34–36], the topological properties of the underlying graph have critical effects on the threshold.

In time-varying networks, the analytical study of contagion processes is hindered by the difficulties in dealing with the concurrent time scales of the contagion and network evolution processes [37–42]. The contagion timescale, τ_P , is defined by the average recovery time, i.e. μ^{-1} . The network timescale, τ_G , is instead dependent on the convolution of the activity time scale, a^{-1} , of each node. Considering these observations, it is possible to derive the mean-field level dynamical equations describing the contagion process. Let us define the activity block variable I_a^t and S_a^t as the number of infected and susceptible individuals, respectively, in the class of activity a at time t . This allows us to write the mean-field evolution of the number of infected individuals in each group of nodes with activity a as

$$I_a^{t+1} = I_a^t - \mu I_a^t + \lambda m (N_a - I_a^t - R_a^t) a \int da' \frac{I_{a'}^t}{N} + (1) \\ + \lambda m (N_a - I_a^t - R_a^t) \int da' \frac{I_{a'}^t a'}{N},$$

where N_a is the total number of individuals with activity rate a (constant over time) and R_a^t are the nodes in the class a at the time t that have been immunized/removed from the network. In Eq. 1, the first term considers the number of infected in class a at time t . The second term describes the number of nodes that recover going back in the class S_a . The third term represents the number of infected individuals generated when nodes in the class $S_a^t = N_a - I_a^t - R_a^t$ are active and connect with infected nodes in the other activity classes. Finally, the last term considers the number of infected generated when nodes in the class S_a^t are linked by active infected nodes in other activity classes. In the absence of any controlling strategy $R_a^t = 0$. In this case, considering the convolution on all activity time scales and ignoring the second order terms in I/N , we can write $I^{t+1} = I^t - t\mu I^t + \lambda m \langle a \rangle I^t + \lambda m \theta^t$, where $\theta^t = \int da' I_{a'}^t a'$. By multiplying both sides of Eq. 1 by a and integrating, we obtain $\theta^{t+1} = \theta^t - \mu \theta^t + \lambda m \langle a^2 \rangle I^t + \lambda m \langle a \rangle \theta^t$. The system equations for I^{t+1} and θ^{t+1} provides an epidemic outbreak only if the dominant eigenvalue of the corresponding matrix is larger than 1. Thus, the epidemic threshold reads as $\frac{\beta}{\mu} \geq \xi^{SIS} \equiv \frac{2\langle a \rangle}{\langle a \rangle + \sqrt{\langle a^2 \rangle}}$, where $\langle a \rangle$ and $\langle a^2 \rangle$ are the first and second moment of

activity distribution [43] (see Ref. [28] and the SI for the calculation details). Remarkably, the threshold does not depend on the time-aggregated network representation. It is defined by the interplay between the time scale of the contagion process and the convolution of the network time scales encoded in the activity distribution moments. In particular, the network's dynamics enter in the expression as the average activity in the system and its fluctuations.

By using this framework, we can study different immunization strategies. First, let us consider the random strategy (RS) in which a fraction p of nodes is immunized with a uniform probability (see Fig. 1-D). In this case, the system of equations describing the dynamic process in activity-driven networks can be obtained by setting $R_a = pN_a$. The epidemic threshold condition changes as

$$\frac{\beta}{\mu} \geq \xi^{RS} \equiv \frac{1}{1-p} \frac{2\langle a \rangle}{\langle a \rangle + \sqrt{\langle a^2 \rangle}} = \frac{\xi^{SIS}}{1-p}. \quad (2)$$

As expected, when a fraction p of nodes is randomly immunized/removed, the epidemic threshold can be written as the threshold with no intervention, ξ^{SIS} , rescaled by the number of nodes still available to the spreading process. Indeed, immunizing/removing random nodes is equivalent to rescaling the per capita spreading rate by the fraction of available nodes $\beta \rightarrow \beta(1-p)$ (see also the SI). Another important quantity is the critical value of immunized/removed nodes, p_c , necessary to halt the contagion process. This quantity is a function of network's structure and the specific features of the contagion process. The explicit value of p_c can be obtained by inverting Eq. 2. In Fig. 2-A, we plot p_c as a function of β/μ keeping fixed the statistical properties underlying network. We then simulate the process for each pair of values and plot the average asymptotic density of infected nodes, I_∞^p , in 10^2 independent realizations. The phase space of the diffusion process is divided into two different regions separated by the red solid line that represents p_c as derived by Eq. 2. In the region below the curve, the spreading process will take over, $p < p_c$. However, in the region above the curve, the fraction of removed/immunized nodes is enough to completely stop the diffusion process, $p \geq p_c$. To further assess the efficiency of the immunization strategy in Fig. 2-D (green triangles), we plot, as a function of the density of removed/immunized nodes p , the ratio I_∞^p/I_∞^0 where I_∞^0 is the asymptotic density of infected nodes when no-intervention is implemented. As shown clearly in the figure, the random strategy allows a reduction in the fraction of infected nodes just for large values of p .

In networks with heavy-tailed degree distributions, targeting nodes with high degree centrality performs more efficiently than random strategies [1, 3]. Analogously,

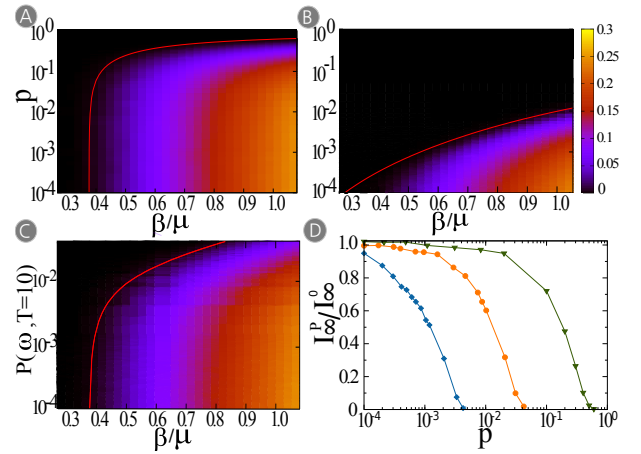


FIG. 2: Panels A, B, and C show the phase space of an SIS process under random, targeted, and egocentric control strategy, respectively. Considering $N = 10^4$, $m = 3$, $\epsilon = 10^{-3}$, activity distributed as $F(a) \sim a^{-2.2}$, we plot I_∞ as a function of β/μ and p . Red curves represent the critical value p_c . Panel D shows the comparison of the stationary state of a SIS model with and without control strategy, I_∞^p/I_∞^0 , as a function of p when $\beta/\mu = 0.81$. In green triangles, we consider the random strategy, in blue diamonds the targeted strategy, and in orange circles the egocentric strategy. Each plot is made averaging 10^2 independent simulations started with 1% of random seeds.

in activity driven networks, effective strategies shall target high activity nodes. For this reason, we rank nodes in decreasing order of activity, immunizing/removing the top ranking pN nodes, see Fig. 2-E. This method is equivalent to fix a value a_c so that any node with activity $a \geq a_c$ is immune to the contagion process [44]. Also, for this scheme, it is possible to derive the analytic expression for the epidemic threshold (see SI):

$$\frac{\beta}{\mu} \geq \xi^{TS} \equiv \frac{2\langle a \rangle}{\langle a \rangle^c + \sqrt{(1-p)\langle a^2 \rangle^c}}, \quad (3)$$

where ξ^{TS} indicates the threshold in the case of the targeted control strategy. In the above expression, we define $\langle a^n \rangle^c = \int_{\epsilon}^{a_c} a^n F(a) da$ as the moments of the activity distribution discounting the removed/immunized nodes. Eq. (3) is not a simple rescaling of the original threshold expression and implies a drastic change in the behavior of the contagion process. As shown in detail in the SI, the analytical predictions (Eq. 3) well reproduce the behavior observed in numerical simulations. In order to define the critical value of p necessary to completely stop the spreading, we have to invert Eq. 3. The moments of the distribution of the remaining nodes are a function of p through a_c . For this reason, it is not possible to derive explicitly p_c . However, it can be easily evaluated numerically by solving the equation $\xi^{TS} - \beta/\mu = 0$ for different values of β/μ . In Fig. 2-B, we show p_c (red line) as a function of β/μ . The efficiency of the targeted strategy is clear. Immunizing/removing a

very small fraction of the most active nodes is enough to stop the contagion process. This is confirmed in Fig. 2-D (blue diamonds) where we plot the ratio I_∞^p/I_∞^0 . The extreme efficiency of this strategy is due to the crucial role of high activity nodes in the spreading process. Immunizing just the top 1% of nodes is enough to halt the disease.

Unfortunately, the network-wide knowledge required to implement targeted control strategies is generally not available [6]. In the case of evolving networks, this issue is even more pronounced as node's characterization depends on how long it is possible to observe the network dynamics. Here we propose a strategy based on the local sampling of the network. A fraction w of randomly selected nodes act as "probes". During an observation time T , we monitor their egonet generated by their interactions in the network. After the observation window, we select randomly a node in the egonet of each probe, immunizing/removing it (see Fig. 1-F). For the sake of comparison with the previous control strategies, we define the fraction of actual immunized nodes as p (in general $w \neq p$ [45]), and the epidemic starts after nodes have been immunized. In this scheme, the probability of immunization/removal for one node with activity a after a time step is:

$$P_a = aw \int da' \frac{mN_{a'}}{N} + w \int da' a' \frac{mN_{a'}}{N} \frac{1}{m}. \quad (4)$$

The first term on the r.h.s. considers the probability that a node of class a is active and reaches one of the probes; the second term, instead, takes into account the probability that one node of class a gets a connection from one active probe. Solving the integrals in Eq. (4), we can write $P_a = w(am + \langle a \rangle)$. Thus, the probability of immunization of one node in the activity class a after t time steps is $P_a^t = 1 - (1 - P_a)^t$, and therefore, summing over all the activity classes, we can estimate the total number of immunized individuals as $R^T = \sum_a N_a P_a^T = \sum_a N_a [1 - (1 - P_a)^T]$. The equation for P_a does not consider the depletion of nodes in each class due to the immunization process. The formulation is then a good approximation for small w and T , when the probability that a probe is selected more than once is very small. Replacing the expression for the removed/immunized individuals in the basic SIS equations (see SI) yields the following epidemic threshold for the egocentric sampling strategy (ESS):

$$\frac{\beta}{\mu} \geq \xi^{ESS} \equiv \frac{2\langle a \rangle}{\Psi_1^T + \sqrt{\Psi_0^T \Psi_2^T}}, \quad (5)$$

where we define $\Psi_n^T = \int da a^n (1 - P_a)^T F(a)$. This last integral is a function of the observing time window T , the probability of immunization/removal of each class and the activity distribution. We evaluate each Ψ term through numerical integration. As shown in the SI, the

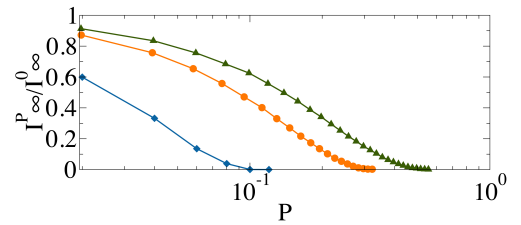


FIG. 3: Comparison of the stationary state of an SIS model with and without control strategies. The I_∞^p/I_∞^0 ratio of the final fraction of diseased nodes of random (egocentric, targeted) strategy I_∞ , and without intervention I_∞^0 is shown with green triangles (orange circles, blue diamonds) as a function of p . Every simulations was initiated with 1% infected seed, executed 10^2 times for $T = 10^4$ step with $\beta/\mu = 2.5$. Each step was integrated for 6×10^2 seconds, and periodic temporal boundary condition was applied.

analytical predictions (Eq. 5) nicely reproduce the behavior observed in numerical simulations. As done for the other two measures, we define the critical value of p by solving numerically the equation $\xi^{ESS} - \beta/\mu = 0$ for different values of β/μ . In Fig. 2-C, we show p_c (red solid line) as a function of β/μ . Based on the plot, it is clear how this strategy is much more efficient than the random one, although not as performant as the targeted scheme (see also Fig. 2-D). The efficiency of this strategy is due to the ability to reach active nodes by a local exploration done observing the systems for few time steps. Real world time-varying networks add a number of complications to the simplified picture offered by activity driven networks. Indeed, they exhibit correlations among nodes, persistency of links, and burstiness of the activity pattern, just to cite a few [8–25]. Therefore, it is extremely important to validate, at least in its basic phenomenology, the above mean-field framework in real world datasets. We consider a mobile phone call data network consisting of 93,190 connected phone users of a single city involved in almost five million calls over 120 days. We simulate the SIS spreading process imposing 1% of randomly chosen infectious nodes and implement the three immunization strategies previously described (see SI for details on the numerical simulations). We report in Fig. 3 the behavior of I_∞^p/I_∞^0 , providing a measure of the effectiveness of each strategy. Remarkably, there is a qualitative agreement between what is observed in real time-varying systems and the analytical results obtained in activity driven networks.

In summary, we have presented an analytical framework that is able to capture the behavior of control strategies for contagion processes in dynamical networks generated by heterogeneous activity distributions. Although the presented theoretical approach deals with a simplified class of time-varying networks, it can potentially extend to more realistic situations and offer a general analytical and interpretative framework

for the understanding of how we can control and mitigate (as well as enhance) contagion phenomena by acting on the basic parameters of the system. Such results find applications in a broad range of real world phenomena ranging from the diffusion of emerging diseases to information and communication processes.

This work has been partially funded by the NSF CCF-1101743 and CMMI-1125095 awards to AV. We thank A.-L. Barabási for the mobile telephone call dataset.

-
- [1] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
- [2] M. Newman, *Networks. An Introduction* (Oxford University Press, 2010).
- [3] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge, 2010).
- [4] C. Butts, *Science* **325**, 414 (2009).
- [5] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **63**, 036104 (2002).
- [6] R. Cohen, S. Havlin, and D. ben Avraham, *Phys Rev. Lett.* **91** (2003).
- [7] M. Kitsak, L. K. Gallos, S. Havlin, and H. A. Makse, *Nature Physics* **6**, 888 (2010).
- [8] M. Morris, *Nature* **365**, 437 (1993).
- [9] M. Morris, *Sexually Transmitted Diseases, K.K. Holmes, et al. Eds.* (McGraw-Hill, 2007).
- [10] A. Clauset and N. Eagle, in *DIMACS Workshop on Computational Methods for Dynamic Interaction Networks* (2007), pp. 1–5.
- [11] A. Vespignani, *Nature Physics* **8**, 32 (2012).
- [12] L. E. C. Rocha, F. Liljeros, and P. Holme, *PLoS Comput Biol* **7**, e1001109 (2011).
- [13] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. V. den Broeck, *J. Theor. Biol* **271**, 166 (2011).
- [14] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck, et al., *BMC Medicine* **9** (2011), URL <http://www.biomedcentral.com/1741-7015/9/87>.
- [15] M. Karsai, M. Kivela, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, *Phys. Rev. E* **83**, 025102 (2011), URL <http://link.aps.org/doi/10.1103/PhysRevE.83.025102>.
- [16] G. Miritello, E. Moro, and R. Lara, *Phys. Rev. E* **83**, 045102 (2011), URL <http://link.aps.org/doi/10.1103/PhysRevE.83.045102>.
- [17] M. Kivela, R. Kumar Pan, K. Kaski, J. Kertesz, J. Saramaki, and M. Karsai (2011), arXiv:1112.4312v1.
- [18] N. Fujiwara, J. Kurths, and A. Díaz-Guilera, *Physical Review E* **83**, 025101 (2011).
- [19] R. Parshani, M. Dickison, R. Cohen, H. E. Stanley, and S. Havlin, *EPL (Europhysics Letters)* **90**, 38004 (2010), URL <http://stacks.iop.org/0295-5075/90/i=3/a=38004>.
- [20] P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza, *PLoS ONE* **6**, e19869 (2011).
- [21] A. Panisson, A. Barrat, C. Cattuto, W. V. den Broeck, G. Ruffo, and R. Schifanello, *Ad Hoc Networks* **10** (2011).
- [22] A. Baronchelli and A. Díaz-Guilera, *Phys. Rev. E* **85**, 016113 (2012).
- [23] M. Starnini, A. Baronchelli, A. Barrat, and R. Pastor-Satorras, *Phys. Rev. E* **85**, 056115 (2012).
- [24] R. Pfitzner, I. Scholtes, A. Garas, C. Tessone, and F. Schweitzer, *Phys. Rev. Lett.* **110**, 19 (2013).
- [25] M. Karsai, N. Perra, and A. Vespignani, arXiv:1303.5966 (2013).
- [26] T. Hoffmann, M. Porter, and R. Lambiotte, *Physical Review E* **86**, 046102 (2012).
- [27] Z. Toroczkai and H. Guclu, *Physica A* **378** (????).
- [28] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, *Scientific Reports* **2**, 469 (2012).
- [29] B. Ribeiro, N. Perra, and A. Baronchelli, *Scientific Reports* **3**, 3006 (2013).
- [30] N. Perra, A. Baronchelli, D. Mocanu, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. Lett.* **109**, 238701 (2012).
- [31] S. Liu, A. Baronchelli, and N. Perra, *Phys. Rev. E* **87** (2013).
- [32] M. Starnini and R. Pastor-Satorras, *Phys. Rev. E* **87**, 062807 (2013).
- [33] M. Keeling and P. Rohani, *Modeling Infectious Disease in Humans and Animals* (Princeton University Press, 2008).
- [34] Y. Wang, D. Chakrabarti, G. Wang, and C. Faloutsos, In *Proc 22nd International Symposium on Reliable Distributed Systems* pp. 25–34 (2003).
- [35] C. Castellano and R. Pastor-Satorras, *Phys. Rev. Lett.* **105**, 218701 (2010).
- [36] R. Durrett, *Proc. Nat. Acad. Sci.* **107**, 4491 (2010).
- [37] B. Prakash, H. Tong, M. Valler, and C. Faloutsos, *Machine Learning and Knowledge Discovery in Databases Lecture Notes in Computer Science* **6323**, 99 (2010).
- [38] M. Starnini, A. Machens, C. Cattuto, A. Barrat, and R. Pastor-Satorras, *Journal of Theoretical Biology* **337**, 89 (2013).
- [39] S. Lee, L. Rocha, F. Liljeros, and P. Holme, *PLoS ONE* **7**, e36439 (2012).
- [40] T. Takaguchi, N. Sato, K. Yano, and N. Masuda, *New J. Phys.* **14**, 093003 (2012).
- [41] J. Tang, C. Mascolo, M. Musolesi, and V. Latora, in *Proceedings of IEEE 12th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM 11)* (2011).
- [42] N. Masuda and P. Holme, *F1000Prime Reports* **5** (2013).
- [43] The activities are defined in the interval $\epsilon \leq a \leq 1$ implying finite moments for any power-law activity distribution, see SI for further details.
- [44] The value of p and a_c are linked by the relation $p = \int_{a_c}^1 F(a) da$
- [45] In order to guarantee that a fraction w of nodes is immunized/removed the systems need to be observed for more than one time step. We define T^* as the average time needed for all the probes to have at least one interaction with other nodes. For any observation time $T < T^*$ the fraction of immunized/removed nodes will be in general $p \leq w$.