

The optimal edge for containing the spreading of SIS model

Jiajun Xian¹, Dan Yang², Liming Pan^{2,3} and Wei Wang^{2,4}

¹ School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, People's Republic of China

² Web Sciences Center, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, People's Republic of China

³ School of Computer Science and Technology, Nanjing Normal University, Nanjing, Jiangsu, 210023, People's Republic of China

⁴ Cybersecurity Research Institute, Sichuan University, Chengdu, 610065, People's Republic of China

E-mail: danyangsjhd@hotmail.com

Received 17 December 2019

Accepted for publication 3 February 2020

Published 14 April 2020



Online at stacks.iop.org/JSTAT/2020/043501
<https://doi.org/10.1088/1742-5468/ab780d>

Abstract. Numerous real-world systems, for instance, communication platforms and transportation systems, can be abstracted into complex networks. Containing spreading dynamics (e.g. epidemic transmission and misinformation propagation) in networked systems is a hot topic on multiple fronts. Most of the previous strategies are based on the immunization of nodes. However, sometimes, these node-based strategies can be impractical. For instance, in train transportation networks, it is excessive to isolate train stations for flu prevention. On the contrary, temporarily suspending some connections between stations is more acceptable. Thus, we pay attention to the edge-based containment strategy. In this study, we develop a theoretical framework to find the optimal edge for containing the spread of the susceptible-infected-susceptible model on complex networks. To be specific, by performing a perturbation method to the discrete-Markovian-chain equations of the SIS model, we derive a formula that approximately provides the decremental outbreak size after the deactivation of a certain edge in the network. Then, we determine the optimal edge by simply choosing the one with the largest decremental outbreak size. Note that our proposed theoretical framework incorporates the information of both network structure and spreading dynamics. Finally, we test the performance of our

method by extensive numerical simulations. Results demonstrate that our strategy always outperforms other strategies that are based only on structural properties (degree or edge betweenness centrality). The theoretical framework in this study can be extended to other spreading models and offers inspiration for further investigations on edge-based immunization strategies.

Keywords: network dynamics

Contents

1. Introduction	2
2. Model description	3
3. Theoretical analysis	4
3.1. The discrete-Markovian-chain approach for the SIS model.....	4
3.2. Determining the optimal edge for containing the spreading.....	5
4. Simulation results	8
5. Conclusions	12
Acknowledgments	13
Appendix	13
A.1. The iteration formula of $\dot{I}(t)$	13
References	15

1. Introduction

The subject of containing spreading dynamics in networked systems has attracted substantial attention from multiple fronts, for instance, network science, statistical physics, and computer science. Some common spreading dynamics, including the epidemic transmission [1, 2] and misinformation spreading [3, 4], can influence all aspects of an individual's life and cause great impact to the socioeconomic systems. The study of containing these spreading dynamics is of both theoretical and practical importance.

Before getting into the problem of spreading containment, it is necessary to build suitable models to describe the spreading dynamics. Researchers have proposed various models for different spreading cases. For instance, the classic susceptible-infected-susceptible (SIS) model [5] and the susceptible-infected-recovered (SIR) model [6, 7], along with many of their extensions [8–10], have been widely applied to describe the spreading of a disease or simple information (e.g. rumors). In these simple contagions, the susceptible individuals could be infected by a single contact with an infected one. When it comes to modeling some complex information spreading, such as the behavior adoption [11] and political information [12, 13], the susceptible individuals first assess the legality of the information and conduct a risk assessment. Then, they become infected with a probability that increases with the cumulative amount of contact with

the infected individuals. This mechanism is referred to as the social reinforcement [14, 15]. The classical threshold model and other models extended from it incorporate this mechanism for complex contagions. More spreading models with other complex mechanisms are discussed in [16].

Based on the different spreading models, researchers go further to develop containing strategies for the spreading dynamics. An effective containing strategy is supposed to effectively increase the spreading outbreak threshold [17] or decrease the outbreak size [18]. Many target containing strategies have been proposed, for instance, immunizing a fraction of nodes according to the centrality indexes of them, like their degree, betweenness, closeness, PageRank, and eigenvector centrality [19–25]. However, in the real-life, the identification of centrality-defined individuals sometimes can be time-consuming. Thus, researchers come up with strategies that do not rely on any centrality indexes of the nodes, such as the acquaintance immunization strategy [19, 21, 26, 26, 27], which is more suitable for practical applications. Besides, many containing strategies inspired by the methods of optimization and control [28–34] have been proposed by researchers as well. There are also some specific containment strategies for networks in different categories, for instance, the multiple networks [35], adaptive networks [36], and temporal networks [37].

All the strategies mentioned above are based on the immunization of nodes. However, these strategies are sometimes difficult to put into practice. For instance, it is excessive to isolate a fraction of train stations within the whole country to prevent the spreading of flu on the train transportation network, but it is more acceptable to suspend some connections in the network. Thus, the study of edge-based containing strategies should be emphasized. Some researchers have proposed strategies of deactivating edges selected by the properties of the adjacent nodes or the edges themselves [38, 39]. Besides, some strategies incorporate both the structural characteristics of the network and parameters of the spreading process [40].

This study focuses on the subject of determining the optimal edge for containing the spreading of the SIS model on complex networks. The theoretical framework we developed can find the optimal or near-optimal edge for the spreading containment of the SIS model. By developing a perturbation method to the discrete-Markovian-chain equations of the SIS model, we obtain a formula that approximately provides the decremental outbreak size after deactivating a certain edge in the network. Then, we determine the optimal edge by simply selecting the one with the largest decremental outbreak size. It is worth mentioning that the information of both network structure and spreading dynamics is considered in our theoretical framework.

The paper is organized as follows. Section 2 provides the model description. The detailed theoretical analysis is presented in section 3. Then, we present the numerical simulations in section 4. Finally, we provide a conclusion in section 5.

2. Model description

In this study, we consider the classic SIS model on a complex network G of N nodes and M edges. The SIS model is extensively applied to describe the spreading of simple

information or disease. Each node in this model can be in two different states, that is, the susceptible state (S) and the infected state (I). Initially, we select a small fraction of nodes to be the infected seeds, keeping the others in the S state. Then, for every time step, each infected node tries to infect neighbors in the S state with probability λ . Afterward, all the nodes in the I state return to the S state with probability γ . Without loss of generality, we set $\gamma = 0.5$ in this study. Eventually, the dynamic system will reach the steady-state on network G , where the fraction of nodes in the I state fluctuates around a certain value ρ , that is, the outbreak size.

Let A be the adjacency matrix of the network G . Thus, A should be a square $N \times N$ matrix such that its element $A_{ij} = 1$ when there is an edge between node i and j , and $A_{ij} = 0$ when there is no edge. Previous studies [41, 42] have demonstrated that the spreading outbreaks when the effective transmission probability $\beta = \lambda/\gamma$ is larger than the reciprocal of the leading eigenvalue ω of A . That is to say, if $\lambda_c/\gamma = 1/\omega$, then the spreading will break out only when λ is larger than the critical value $\lambda_c = \gamma/\omega$. Otherwise, if $\lambda < \lambda_c$, then no outbreaks will be observed, and no containment process is needed. Therefore, we focus on the case when $\lambda > \lambda_c$ in this study.

After deactivating a specific edge (i_0, j_0) in the networks G , we get a new network G' . And the adjacency matrix A' of G' should be

$$A' = A - \dot{A}, \quad (1)$$

where the element $\dot{A}_{ij} = 1$ only when $(i, j) \in \{(i_0, j_0), (j_0, i_0)\}$. Denote ρ' as the outbreak size on the new network G' . We aim to find the optimal edge, which, if deactivated, can maximize the decremental outbreak size $\dot{\rho} = \rho - \rho'$.

3. Theoretical analysis

In this section, we first present the discrete-Markovian-chain (DMC) approach [42, 43] for the SIS model on the network G . Then, using a perturbation method for the DMC, we derive a formula that approximately provides the decremental outbreak size after deactivating an edge in the network G . Finally, using the formula, we study the problem of determining the optimal edge, which, if deactivated, can maximize the decremental outbreak size.

3.1. The discrete-Markovian-chain approach for the SIS model

In this subsection, we adopt the discrete-Markovian-chain (DMC) approach to study the SIS model on the network G . The DMC approach can accurately predict the phase diagram for contact-based spreading dynamics in complex networks and overcomes the computational cost of Monte Carlo simulations [42].

To begin with, we define a set of discrete-time equations for the probability of individual nodes to be infected. Denote $I_i(t)$ as the probability that node i is in the I state at time t . Then, the node i is in the S state at time t with probability $S_i(t) = 1 - I_i(t)$. If i is in the I state at $t + 1$, then either it was in the I state at t and has not recovered,

or it was in the S state at t and has been infected by its infected neighbors. Thus, the evolution of $I_i(t)$ is

$$I_i(t+1) = (1-\gamma)I_i(t) + [1-\Theta_i(t)]S_i(t), \quad (2)$$

where $1-\Theta_i(t)$ is the probability that node i gets infected at time t , and

$$\Theta_i(t) = \prod_{j=1}^N [1 - \lambda A_{ij} I_j(t)]. \quad (3)$$

When the dynamic system reaches the steady state, we have $I_i(t) = I_i(t+1) = \tilde{I}_i$, $S_i(t) = S_i(t+1) = \tilde{S}_i$, and $\Theta_i(t) = \Theta_i(t+1) = \tilde{\Theta}_i$. Taking all the nodes into consideration, the equations (2) and (3) can be written in terms of vectors in the steady state as

$$\tilde{I} = (1-\gamma)\tilde{I} + (1-\tilde{\Theta}) \circ \tilde{S}, \quad (4)$$

and

$$\tilde{\Theta}_i = \prod_{j=1}^N (1 - \lambda A_{ij} \tilde{I}_j), \quad (5)$$

where \tilde{I} and $\tilde{\Theta}$ are vectors of length N with entries $\tilde{I} = (\tilde{I}_1, \dots, \tilde{I}_N)^T$ and $\tilde{\Theta} = (\tilde{\Theta}_1, \dots, \tilde{\Theta}_N)^T$, respectively, and \circ denotes component-wise vector product. Combining the equations (4) and (5), we obtain the expected outbreak size of the spreading on networks G as follows:

$$\rho = N^{-1} \mathbf{1}^T \tilde{I}. \quad (6)$$

3.2. Determining the optimal edge for containing the spreading

For convenience, we denote the new network we get after deactivating the specific edge l in the original network by G'_l . Besides, the decremental outbreak size $\dot{\rho}$ of the SIS model on network G after deactivating the specific edge l is denoted by $\dot{\rho}_l$. The spreading outbreak size ρ_l on the new network G_l depends on the position of the edge l . In this subsection, we introduce a perturbation method to obtain an approximate estimate of the decremental outbreak size $\dot{\rho} = \rho - \rho_l$ after the deactivation of edge l . Then, we use the approximate estimate to determine the optimal edge, which, if deactivated, can maximize the decremental outbreak size $\dot{\rho}$.

Inferring from equation (1), we get the DMC equations of network G_l as follows:

$$I_i(t+1) = (1-\gamma)I_i(t) + [1-\Theta_i(t)][1-I_i(t)], \quad (7)$$

and

$$\Theta_i(t) = \prod_{j=1}^N [1 - \lambda(A_{ij} - \dot{A}_{ij})I_j(t)]. \quad (8)$$

On the consideration that the fixed point $I(\infty) = [I_1(\infty), \dots, I_N(\infty)]^T$ of G_l will stay close to the fixed point \tilde{I} of G since only one edge has been deactivated, we iterate equations (7) and (8) with initial condition $I(0) = \tilde{I}$. Then, we employ the decomposition of

$I(t) = \tilde{I} + \dot{I}(t)$ and $\Theta(t) = \tilde{\Theta} + \dot{\Theta}(t)$. According to appendix, we can obtain the iteration formula of $\dot{I}(t)$ as

$$\begin{aligned} \dot{I}(t+1) = & (\tilde{\Theta} - \gamma)\dot{I}(t) + (1 - \tilde{I}) \circ \lambda \tilde{\Theta} \circ (A - \dot{A}) \Psi \dot{I}(t) \\ & + (1 - \tilde{I}) \circ \tilde{\Theta} \circ \dot{A} \log(1 - \lambda \tilde{I}), \end{aligned} \quad (9)$$

where Ψ is the $N \times N$ diagonal matrix with entries $\Psi_{ij} = 1/(1 - \lambda \tilde{I}_j)$ for $i = j$ and $\Psi_{ij} = 0$ for $i \neq j$. This equation can be written in terms of matrix multiplication as follows:

$$\dot{I}(t+1) = \Xi \dot{I}(t) + \xi, \quad (10)$$

where

$$\Xi = \lambda \text{diag}(\tilde{\Theta} - \tilde{I} \circ \tilde{\Theta})(A - \dot{A})\Psi + \text{diag}(\tilde{\Theta} - \gamma) \quad (11)$$

and

$$\xi = (1 - \tilde{I}) \circ \tilde{\Theta} \circ \dot{A} \log(1 - \lambda \tilde{I}). \quad (12)$$

Here, $\text{diag}(\cdot)$ denotes the diagonal matrix with the elements of the input vector as diagonal entries. Thus, the stationary solution $\dot{I}(\infty)$ of the perturbed system satisfies

$$\dot{I}(\infty) = \Xi \dot{I}(\infty) + \xi, \quad (13)$$

or in the closed form

$$\dot{I}(\infty) = (\mathbb{I} - \Xi)^{-1} \xi. \quad (14)$$

Thus, an explicit relationship between the deactivated edge l and the decremental outbreak size can be found as

$$\dot{\rho}_l = -N^{-1} \mathbf{1}^T \dot{I}(\infty) = -N^{-1} \mathbf{1}^T (\mathbb{I} - \Xi)^{-1} \xi. \quad (15)$$

In order to solve equation (15), we decompose Ξ as $\Xi = \Xi^0 + \dot{\Xi}$, where

$$\Xi^0 = \lambda \text{diag}(\tilde{\Theta} - \tilde{I} \circ \tilde{\Theta}) A \Psi + \text{diag}(\tilde{\Theta} - \gamma) \quad (16)$$

depends only on A , and

$$\dot{\Xi} = -\lambda \text{diag}(\tilde{\Theta} - \tilde{I} \circ \tilde{\Theta}) \dot{A} \Psi \quad (17)$$

depends only on \dot{A} . For any edge $l = (i, j)$, the matrix \dot{A} can be written as sum of two outer products

$$\dot{A} = uv^T + vu^T, \quad (18)$$

where u, v are vectors of length N with $u_k = \delta_{k,i}$ and $v_k = \delta_{k,j}$ for $1 \leq k \leq N$. Define short notations as

$$\varepsilon_{ij} := -\lambda \left(\tilde{\Theta}_i - \tilde{I}_i \tilde{\Theta}_i \right) \left(1 - \lambda \tilde{I}_j \right)^{-1}, \quad (19)$$

then it is easy to check that

$$\dot{\Xi} = \varepsilon_{ij}uv^T + \varepsilon_{ji}vu^T. \quad (20)$$

The Sherman–Morrison formula says that

$$(\mathbb{I} - \Xi)^{-1} = (\mathbb{I} - \Xi^0 - \varepsilon_{ij}uv^T - \varepsilon_{ji}vu^T)^{-1} = X + \frac{\varepsilon_{ij}Xuv^TX}{1 - \varepsilon_{ij}X_{ji}}, \quad (21)$$

where

$$X = (\mathbb{I} - \Xi^0 - \varepsilon_{ij}uv^T)^{-1}. \quad (22)$$

Applying the Sherman–Morrison formula again gives

$$X = Y + \frac{\varepsilon_{ji}Yvu^TY}{1 - \varepsilon_{ji}Y_{ij}}, \quad (23)$$

where

$$Y = (\mathbb{I} - \Xi^0)^{-1}. \quad (24)$$

Again we define short notations for convenience as follows,

$$c_{ij} = (\tilde{\Theta}_i - \tilde{I}_i\tilde{\Theta}_i) \log(1 - \lambda\tilde{I}_j). \quad (25)$$

Then ξ can be checked satisfying

$$\xi = c_{ij}u + c_{ji}v. \quad (26)$$

By substituting equations (21) and (26) into equation (15), we can get

$$\begin{aligned} \dot{\rho}_l &= c_{ij}1^TXu + c_{ji}1^TXv + \frac{\varepsilon_{ij}c_{ij}X_{ji}1^TXu}{1 - \varepsilon_{ij}X_{ji}} \\ &\quad + \frac{\varepsilon_{ij}c_{ji}X_{jj}1^TXu}{1 - \varepsilon_{ij}X_{ji}}. \end{aligned} \quad (27)$$

According to equation (23), we carefully expand equation (27) and get

$$\begin{aligned} \dot{\rho}_l &= -N^{-1} \left[\frac{(c_{ij} - c_{ij}\varepsilon_{ji}Y_{ij} + c_{ji}\varepsilon_{ij}Y_{jj})1^TYu}{(1 - \varepsilon_{ij}Y_{ji})(1 - \varepsilon_{ji}Y_{ij}) - \varepsilon_{ij}\varepsilon_{ji}Y_{ii}Y_{jj}} \right. \\ &\quad \left. + \frac{(c_{ji} - c_{ji}\varepsilon_{ij}Y_{ji} + c_{ij}\varepsilon_{ji}Y_{ii})1^TYv}{(1 - \varepsilon_{ij}Y_{ji})(1 - \varepsilon_{ji}Y_{ij}) - \varepsilon_{ij}\varepsilon_{ji}Y_{ii}Y_{jj}} \right]. \end{aligned} \quad (28)$$

Equation (28) provides the approximate decremental outbreak size after deactivating the edge $l = (i, j)$ in the network G . We can use the formula to determine the optimal edge for containing the spreading of the SIS model by simply selecting the edge with the highest $\dot{\rho}_l$. As one can see, we obtain equation (28) through complicated derivations. Looking at the right-hand of equation (28), we can observe that it incorporates both the information of networks structure (i.e. the adjacency matrix A) and spreading dynamics (i.e. λ , γ , \tilde{I} and $\tilde{\Theta}$). Section 4 will show that equation (28) gives a good prediction of the decremental outbreak size.

4. Simulation results

For convenience, we refer to the containing strategy of deactivating the optimal edge L selected by equation (28) as the perturbation-based-strategy (PBS). In this section, extensive numerical simulations are performed to verify the containing performance of the PBS. Both synthetic and real-world networks are considered in our simulations. Note that the DMC approach can predict the results of the Monte Carlo simulations accurately and have a lower computational cost; thus, we conduct our numerical simulations based on the DMC approach instead of the Monte Carlo method [42].

According to section 3, the proposed PBS in this study incorporates the information of both network structure and spreading dynamics. To better understand the importance of dynamic information in the PBS, we employ two contrasting strategies that only consider the structure of networks. Denote f^b as the edge betweenness centrality. Besides, the edge betweenness of the specific edge $l = (i, j)$ is denoted by f_l^b . The first contrast strategy is to deactivate the edge with the highest f^b . We refer to this strategy as betweenness-centrality-strategy (BCS) and the specific edge selected by the BCS as L^B . Similarly, the second contrast strategy is based on the degree k of nodes; thus, it can be referred to as degree-based-strategy (DBS). Specifically, for the DBS, we select the edge with the highest degree product f^d . For the specific edge $l = (i, j)$, the degree product should be $f_l^d = k_i k_j$. The edge selected by the DBS is denoted by L^D .

First, we perform the containing strategies on two synthetic networks G^1 and G^2 . Both of them are scale-free (SF) networks with degree distribution $p(k) \sim k^{-\alpha}$, where α denotes the degree exponent. We set $\alpha_1 = 2.3$ and $\alpha_2 = 3.0$ as the degree exponent of network G^1 and G^2 , respectively. One can find more information about the two networks in table 1. Denote $\hat{\rho}$ as the decremental outbreak size obtained by simulations after deactivating the selected edge. Then, we rank the edges according to the values of $\hat{\rho}$. We refer to this kind of edge rank as numerical rank r in the rest of the paper. To compare the performance of the strategies, we are particularly concerned about the optimal edges L , L^B , and L^D selected by PBS, BCS, and DBS, respectively. We compute the normalized numerical rank $R = r/M$ of edges L , L^B and L^D . The smaller the R , the better the performance. As shown in the figures 1(a) and (b), the DBS performs well on both synthetic networks when λ near the critical value λ_c , but the performance fails quickly when λ becomes large. The BCS performs well only for several values of λ . However, the PBS performs well on both networks for all the values of λ . Figures 1(c) and (d) show the corresponding $\hat{\rho}$ of edges L , L^B and L^D on the synthetic networks G^1 and G^2 , respectively. The larger the $\hat{\rho}$, the better the performance. By comparing the $\hat{\rho}$ of L , L^B and L^D , we can draw the same conclusion as that demonstrated in figures 1(a) and (b).

Second, we investigate the overall correlations between the edge ranks scored by equation (28) and the numerical ranks. To begin with, we compare the decremental outbreak size $\hat{\rho}$ numerically computed by equation (28) with the decremental outbreak size $\hat{\rho}$ obtained by simulations. Setting $\lambda = 0.1$, the results of $\hat{\rho}$ versus $\hat{\rho}$ on the synthetic networks G^1 and G^2 are shown in the figures 2(a) and (b), respectively. The results demonstrate that the values of $\hat{\rho}$ and $\hat{\rho}$ are almost linearly correlated; that is to say, equation (28) can predict the decremental outbreak size well. To better understand the rank correlations for all the values of λ , we employ the Spearman rank correlation

Table 1. Basic statistics of the two synthetic networks and six real-world networks employed in this study: the number of nodes N , the number of edges M , the average degree $\langle k \rangle$, and the theoretical spreading threshold λ_c .

Name	N	M	$\langle k \rangle$	λ_c
SF2.3	200	1000	10	0.076
SF3.0	200	1000	10	0.083
Residence hall	217	1839	16.949	0.046
Hamsterster friendships	1788	12 476	13.955	0.022
Jazz musicians	198	2742	27.697	0.025
Facebook (NIPS)	2888	2981	2.0644	0.036
Physicians	117	465	7.95	0.099
Air traffic control	1226	2408	3.928	0.109

coefficient [44, 45] to quantify the mentioned correlations. The Spearman rank correlation coefficient is defined as

$$m_s = 1 - 6 \frac{\sum_{l=1}^M (\hat{r}_l - \dot{r}_l)^2}{M(M^2 - 1)}, \quad (29)$$

where \hat{r}_l and \dot{r}_l are the ranks of edge l scored by $\hat{\rho}$ and $\dot{\rho}$, respectively. Figures 3(a) and (b) shows the results of m_s versus λ on the synthetic networks G^1 and G^2 , respectively. It can be seen that the value of m_s stays close to one for all the values of λ on both networks. That is to say, the edge ranks predicted by the equation (28) and the numerical ranks are strongly correlated for all the values of λ . Similarly, we also compute the Spearman rank correlation coefficient between the edge ranks scored by f^b and the numerical ranks, along with the Spearman rank correlation coefficient between the edge ranks scored by f^d and the numerical ranks. As shown in figures 3(a) and (b), the edge ranks scored by f^b or f^d are positively correlated with the numerical ranks only when λ is near the λ_c . When λ becomes large, their correlations become negative. The results in figures 3(a) and (b) demonstrate that the equation (28) is sufficient to predict the numerical rank of edges, but the f^b or f^d is far from sufficient.

We now go further to investigate the structural properties of the optimal edge L selected by equation (28). The normalized structural statistics f_L^b/f_{\max}^b , f_L^c/f_{\max}^c , f_L^d/f_{\max}^d and f_L^e/f_{\max}^e versus λ are shown in figures 4(a)–(d), respectively, where f_L^x (f_L^e) denotes the product of the closeness centrality (eigenvector centrality) of the nodes at the two ends of edge L . Note that f_{\max}^x is the maximum value in $\{f_l^x\}$, where $1 \leq l \leq M$ and $x \in \{b, c, d, e\}$. As demonstrated in figure 4, when the value of λ is small, the optimal edge L has large f_L^b/f_{\max}^b , f_L^c/f_{\max}^c , f_L^d/f_{\max}^d , and f_L^e/f_{\max}^e . That is to say, when λ is slightly above the critical point λ_c , the optimal edges should be those have high edge betweenness centrality and connect nodes with high closeness, degree, and eigenvector centrality. This can be explained by the fact that the outbreak size is small near the critical point, and the edge between nodes with high centrality can help to keep the cluster of nodes in I state. Thus, deactivating the edge with high centrality can contain the spread well when λ is small. However, when λ becomes large, the optimal edge L will have low f_L^d/f_{\max}^d , f_L^e/f_{\max}^e and middle f_L^b/f_{\max}^b , f_L^c/f_{\max}^c . This is because nodes with high centrality will have a high probability of being infected when λ is large; thus, deactivating the edge between nodes with a high centrality becomes unnecessary.

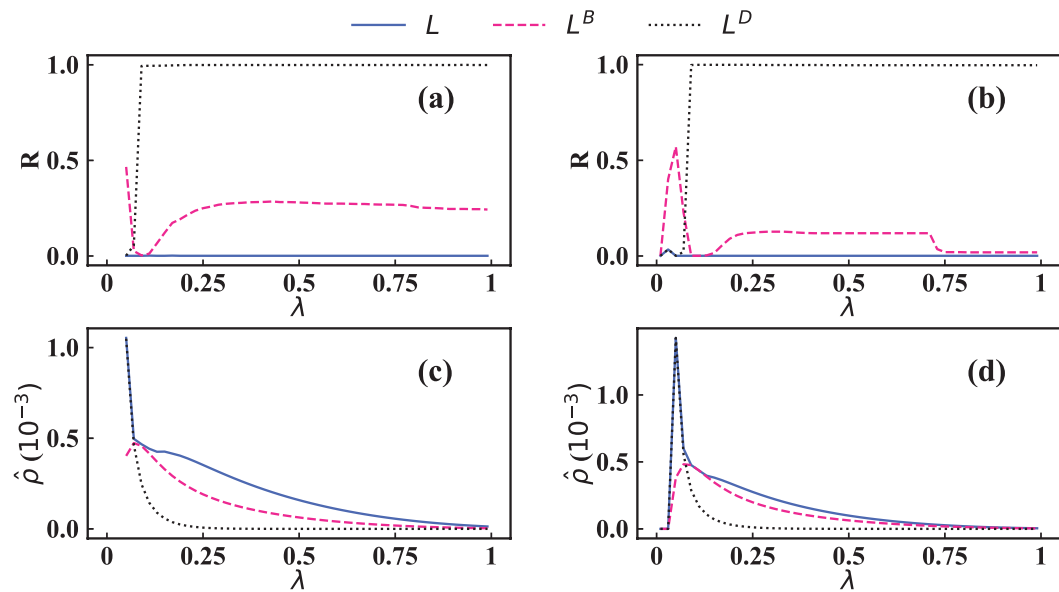


Figure 1. Performance of different strategies versus λ on synthetic networks. The normalized numerical rank R of the optimal edges L (blue solid line), L^B (red dashed line), and L^D (black dotted line) on the SF networks with (a) $\alpha_1 = 2.3$, and (b) $\alpha_2 = 3.0$. The decremental outbreak size $\hat{\rho}$ obtained by simulations after deactivating the optimal edge L (blue solid line), L^B (red dashed line), and L^D (black dotted line) on the SF networks with (c) $\alpha_1 = 2.3$, and (d) $\alpha_1 = 3.0$. More information about the two synthetic networks can be found in table 1.

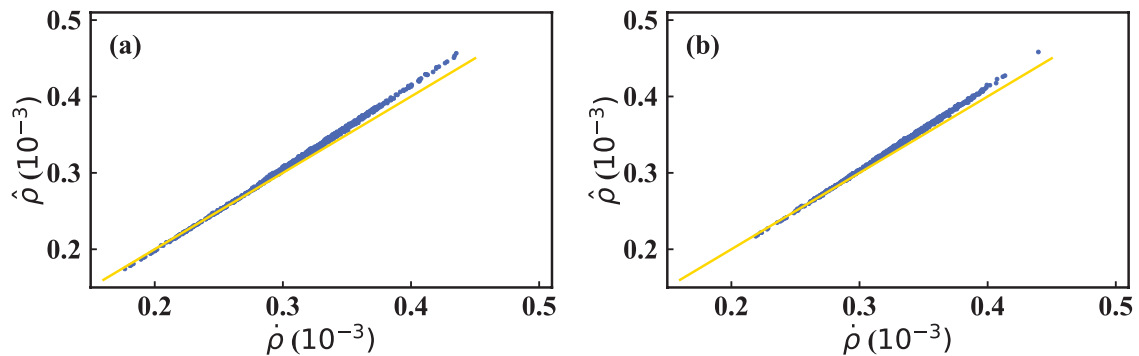


Figure 2. The correlation between decremental outbreak size $\hat{\rho}$ and $\dot{\rho}$. The correlation between decremental outbreak size $\dot{\rho}$ numerically computed by equation (28) and the decremental outbreak size $\hat{\rho}$ obtained by simulations on the SF networks with (a) $\alpha_1 = 2.3$, and (b) $\alpha_2 = 3.0$. The dynamical parameters are set to be $\lambda = 0.1$ and $\gamma = 0.5$. The yellow solid lines in the plots represent the function $\hat{\rho} = \dot{\rho}$.

Finally, we test the performance of the strategies on six real-world networks: (a) Residence hall [46], (b) Jazz musicians [47], (c) Facebook (NIPS) [48], (d) Air traffic control [50], (e) Hamsterster friendships [50], and (f) Physicians [49]. Table 1 provides some basic statistics of these networks. More detailed information on these networks can be found in [50], where they are downloaded from. Figures 5(a)–(i) show the

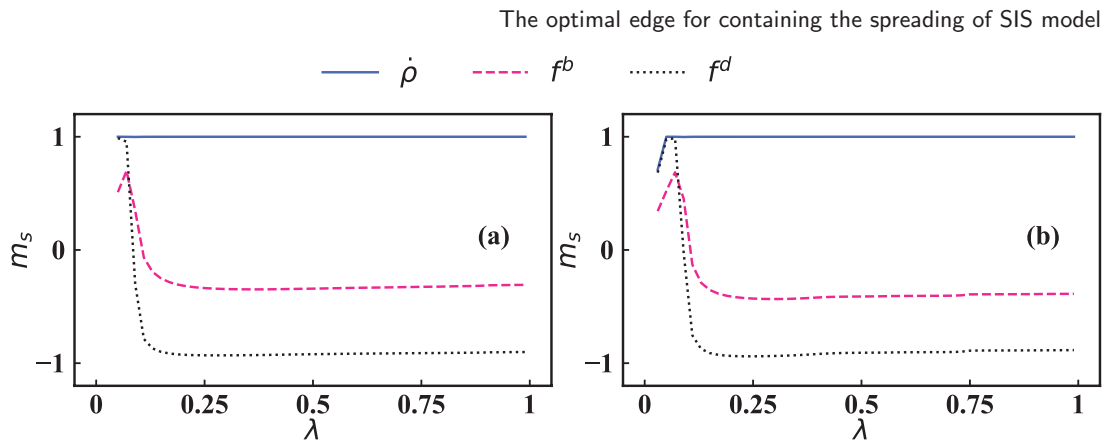


Figure 3. The correlations between the approximate edge ranks and the numerical edge ranks. The Spearman's rank correlation coefficient m_s between the edge ranks scored by equation (28) (blue solid line) and the numerical ranks versus λ on the SF networks with (a) $\alpha_1 = 2.3$, and (b) $\alpha_2 = 3.0$. The Spearman's rank correlation coefficient between the edge ranks scored by the edge betweenness centrality f^b and the numerical ranks are denoted by red dashed lines. Black dotted lines denote the Spearman's rank correlation coefficient between the edge ranks scored by degree product f^d and the numerical ranks. More information about the two synthetic networks can be found in table 1.

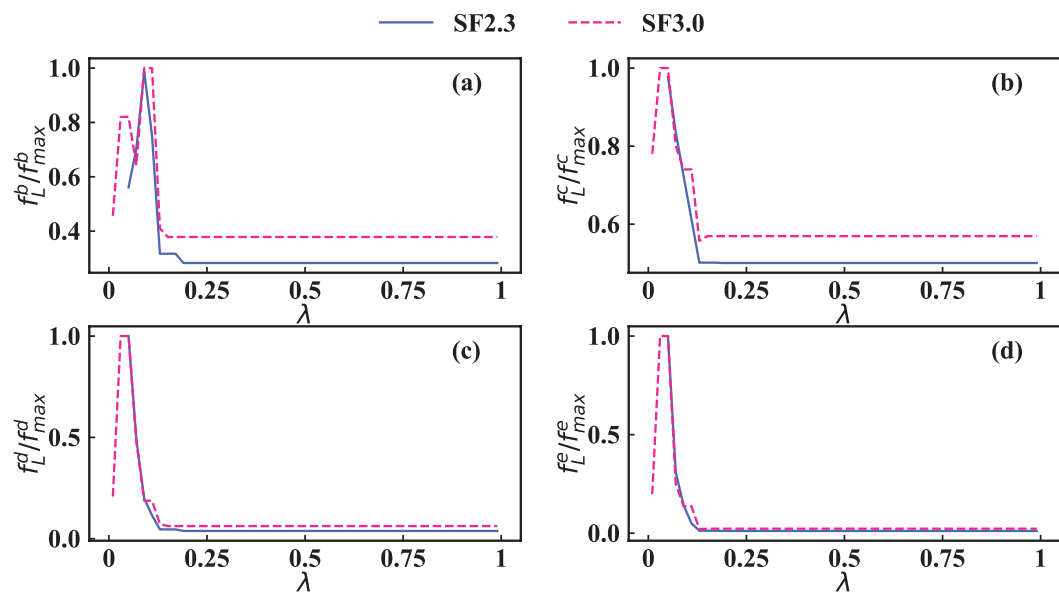


Figure 4. Normalized structural properties of the optimal edge L selected by equation (28). (a) f_L^b/f_{\max}^b versus λ , that is, the normalized edge betweenness centrality of L . (b) f_L^c/f_{\max}^c versus λ , that is, the normalized product of the closeness centrality of nodes at the two ends of L . (c) f_L^d/f_{\max}^d versus λ , that is, the normalized product of the degree of nodes at the two ends of L . (d) f_L^e/f_{\max}^e versus λ , that is, the normalized product of the eigenvector centrality of nodes at the two ends of L . Blue solid lines (red dashed lines) denotes the corresponding results of the SF network with $\alpha_1 = 2.3$ ($\alpha_2 = 3.0$). Some structural properties about the two synthetic networks can be found in table 1.

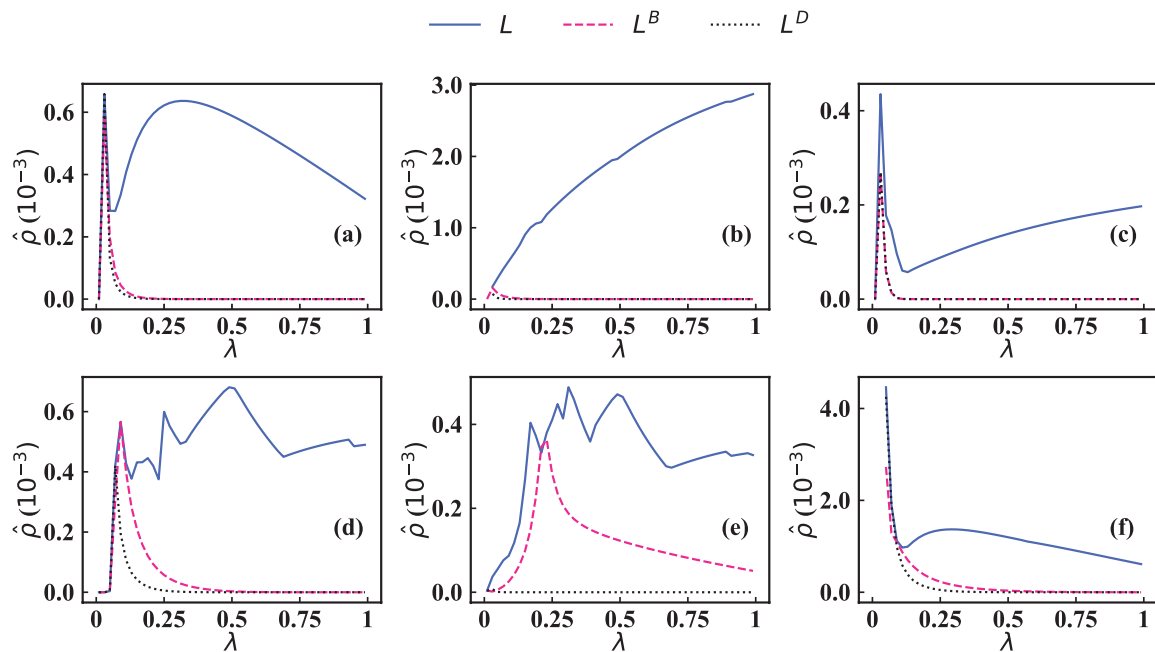


Figure 5. Performance of different strategies on real-world networks. The decremental outbreak size $\hat{\rho}$ obtained by simulations after deactivating the optimal edge L (blue solid line), L^B (red dashed line), and L^D (black dotted line) on the real-world networks (a) Hall of residence, (b) Jazz musicians, (c) Facebook (NIPS), (d) Air traffic control, (e) Hamsterster friendships, and (f) Physicians. Detailed information about the six real-word networks can be found in table 1.

decremental outbreak size $\hat{\rho}$ obtained by simulations after deactivating the optimal edges L , L^B or L^D on the six real-world networks for different transmission probability λ . The results demonstrate that the PBS outperforms the two contrast strategies on all the six real-networks and for all the values of λ studied.

5. Conclusions

Containing spreading dynamics (e.g. epidemic transmission and misinformation propagation) in the networked systems (e.g. transportation systems and communication platforms) is of both theoretical and practical importance. In this study, we developed a theoretical framework to determine the optimal edge for containing the spreading of the SIS model in complex networks.

To be specific, we performed a perturbation method to the DMC equations of the SIS model and obtained a formula that provides an approximate value of the decremental outbreak size after deactivating a certain edge in the network. Afterward, we determined the optimal edge by selecting the one with the largest decremental outbreak size. It is worth mentioning that the formula we obtained incorporates the information of both network structure and spreading dynamics. Extensive numerical simulations

on both synthetic networks and real-world networks demonstrated that our strategy performs well for all the values of λ and outperforms those strategies based only on structure statistics (degree or edge betweenness centrality).

Previous strategies of containing spreading dynamics on complex networks are mostly based on node immunization, which in practice can sometimes be socially and politically difficult. The theoretical framework developed in this study offers inspiration for investigations on edge-based immunization strategies, which could be more practical for some specific real situations. Our theoretical framework could also be extended to other spreading models.

Acknowledgments

This work was partially supported by the China Postdoctoral Science Special Foundation (Grant No.~2019T120829), National Natural Science Foundation of China (Grant Nos.~61903266 and 61603074), Fundamental Research Funds for the Central Universities, and Sichuan Science and Technology Program (NO. 20YYJC4001).

Appendix

A.1. The iteration formula of $\dot{I}(t)$

This appendix shows the detailed steps of obtaining the iteration formula of $\dot{I}(t)$. Based on the decomposition of $I(t) = \tilde{I} + \dot{I}(t)$ and $\Theta(t) = \tilde{\Theta} + \dot{\Theta}(t)$, we get

$$\tilde{I} + \dot{I}(t+1) = (1-\gamma)(\tilde{I} + \dot{I}(t)) + (1-\tilde{I} - \dot{I}(t)) \circ (1-\tilde{\Theta} - \dot{\Theta}(t)), \quad (\text{A.1})$$

where $\dot{I}(t)$ and $\dot{\Theta}(t)$ are assumed small. Ignoring the second-order term $\dot{I}(t) \circ \dot{\Theta}(t)$, we can get

$$\dot{I}(t+1) = (\tilde{\Theta} - \gamma)\dot{I}(t) - (1-\tilde{I}) \circ \dot{\Theta}(t) \quad (\text{A.2})$$

by expanding equation (A.1) and substituting $\tilde{I} = (1-\gamma)\tilde{I} + (1-\tilde{\Theta}) \circ \tilde{S}$. Similarly, $\Theta(t)$ becomes

$$\tilde{\Theta}_i + \dot{\Theta}_i(t) = \prod_{j=1}^N \{1 - \lambda(A_{ij} - \dot{A}_{ij})[\tilde{I}_j + \dot{I}_j(t)]\}. \quad (\text{A.3})$$

We notice the following equation holds

$$[1 - \lambda(A_{ij} - \dot{A}_{ij})][\tilde{I}_j + \dot{I}_j(t)] = \frac{1 - \lambda A_{ij}[\tilde{I}_j + \dot{I}_j(t)]}{1 - \lambda \dot{A}_{ij}[\tilde{I}_j + \dot{I}_j(t)]}, \quad (\text{A.4})$$

which can be checked by substituting all possible combinations of A_{ij}^0 and \dot{A}_{ij} . Divide by $\tilde{\Theta}_i$ for both sides of equation (A.3) and substitute $\tilde{\Theta}_i = \prod_{j=1}^N (1 - \lambda A_{ij} \tilde{I}_j)$ gives

$$1 + \frac{\dot{\Theta}_i(t)}{\tilde{\Theta}_i} = \prod_{j=1}^N \left(1 - \frac{\lambda A_{ij} \dot{I}_j(t)}{1 - \lambda A_{ij} \tilde{I}_j} \right) \times \prod_{j=1}^N \left(1 - \frac{\lambda \dot{A}_{ij} \dot{I}_j(t)}{1 - \lambda \dot{A}_{ij} \tilde{I}_j} \right)^{-1} \\ \times \prod_{j=1}^N \left(1 - \lambda \dot{A}_{ij} \tilde{I}_j \right)^{-1}. \quad (\text{A.5})$$

Note that the following relation holds

$$\frac{\lambda A_{ij} \dot{I}_j(t)}{1 - \lambda A_{ij} \tilde{I}_j} = A_{ij} \frac{\lambda \dot{I}_j(t)}{1 - \lambda \tilde{I}_j}, \quad (\text{A.6})$$

since $A_{ij} \in \{0, 1\}$. Similarly, when replacing A_{ij} in equation (A.6) by $\dot{A}_{ij} \in \{0, 1\}$, we get

$$\frac{\lambda \dot{A}_{ij} \dot{I}_j(t)}{1 - \lambda \dot{A}_{ij} \tilde{I}_j} = \dot{A}_{ij} \frac{\lambda \dot{I}_j(t)}{1 - \lambda \tilde{I}_j}. \quad (\text{A.7})$$

Taking the logarithm on both sides of equation (A.5), expanding to the first orders of $\delta p_i(t)$, $\delta q_i(t)$, and applying the above relation can give

$$\frac{\dot{\Theta}_i(t)}{\tilde{\Theta}_i} = - \sum_{j=1}^N A_{ij} \frac{\lambda \dot{I}_j(t)}{1 - \lambda \tilde{I}_j} + \sum_{j=1}^N \dot{A}_{ij} \frac{\lambda \dot{I}_j(t)}{1 - \lambda \tilde{I}_j} - \sum_{j=1}^N \log \left(1 - \lambda \dot{A}_{ij} \tilde{I}_j \right). \quad (\text{A.8})$$

The terms in the last summation can be checked to satisfy

$$\log \left(1 - \lambda \dot{A}_{ij} \tilde{I}_j \right) = \dot{A}_{ij} \log \left(1 - \lambda \tilde{I}_j \right). \quad (\text{A.9})$$

With the above calculations, equation (A.8) can be written in the matrix form as

$$\dot{\Theta}(t) = -\lambda \tilde{\Theta} \circ \left(A - \dot{A} \right) \Psi \dot{A}(t) - \tilde{\Theta} \circ \dot{A} \log(1 - \lambda \tilde{I}), \quad (\text{A.10})$$

where $\log(1 - \lambda \tilde{I})$ is the vector obtained by taking the logarithm in each entry of $1 - \lambda \tilde{I}$. And Ψ is the $N \times N$ diagonal matrix with entries

$$\Psi_{ij} = \delta_{ij} \frac{1}{1 - \lambda \tilde{I}_j}, \quad (\text{A.11})$$

where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}. \quad (\text{A.12})$$

Substituting equation (A.10) back into equation (A.2) yields the following iteration formula for $\dot{I}(t)$:

$$\dot{I}(t+1) = (\tilde{\Theta} - \gamma) \dot{A}(t) + (1 - \tilde{I}) \circ \lambda \tilde{\Theta} \circ \left(A - \dot{A} \right) \Psi \dot{A}(t) \\ + (1 - \tilde{I}) \circ \tilde{\Theta} \circ \dot{A} \log(1 - \lambda \tilde{I}). \quad (\text{A.13})$$

References

- [1] Wang Z, Moreno Y, Boccaletti S and Perc M 2017 *Chaos, Solitons Fractals* **103** 177–83
- [2] Wang L and Wu J T 2018 *Nat. Commun.* **9** 218
- [3] Xian J, Yang D, Pan L, Wang W and Wang Z 2019 *Chaos* **29** 113123
- [4] Wang W, Ma Y, Wu T, Dai Y, Chen X and Braunstein L A 2019 *Chaos* **29** 123131
- [5] Fu X, Small M, Walker D M and Zhang H 2008 *Phys. Rev. E* **77** 036113
- [6] Daley D J and Kendall D G 1964 *Nature* **204** 1118
- [7] Daley D J and Kendall D G 1965 *IMA J. Appl. Math.* **1** 42–55
- [8] Zuzek L A, Stanley H and Braunstein L 2015 *Sci. Rep.* **5** 12151
- [9] Wang Z, Guo Q, Sun S and Xia C 2019 *Appl. Math. Comput.* **349** 134–47
- [10] Xia C, Wang Z, Zheng C, Guo Q, Shi Y, Dehmer M and Chen Z 2019 *Inf. Sci.* **471** 185–200
- [11] Aral S and Nicolaides C 2017 *Nat. Commun.* **8** 14753
- [12] Romero D M, Meeder B and Kleinberg J 2011 Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter *Proc. 20th Int. Conf. on World Wide Web (ACM)* pp 695–704
- [13] Centola D and Macy M 2007 *Am. J. Sociol.* **113** 702–34
- [14] Watts D J 2002 *Proc. Natl Acad. Sci. USA* **99** 5766–71
- [15] Lü L, Chen D B and Zhou T 2011 *New J. Phys.* **13** 123005
- [16] Wang W, Liu Q H, Liang J, Hu Y and Zhou T 2019 *Phys. Rep.* **820** 1–51
- [17] Matsuki A and Tanaka G 2019 *Phys. Rev. E* **100** 022302
- [18] Xian J, Yang D, Pan L, Liu M and Wang W 2020 *J. Stat. Mech.* 023402
- [19] Wang Z, Bauch C T, Bhattacharyya S, d’Onofrio A, Manfredi P, Perc M, Perra N, Salathe M and Zhao D 2016 *Phys. Rep.* **664** 1–113
- [20] Holme P, Kim B J, Yoon C N and Han S K 2002 *Phys. Rev. E* **65** 056109
- [21] Cohen R, Havlin S and Ben-Avraham D 2003 *Phys. Rev. Lett.* **91** 247901
- [22] Chung F, Horn P and Tsias A 2009 *Internet Math.* **6** 237–54
- [23] Prakash B A, Adamic L, Iwashyna T, Tong H and Faloutsos C 2013 Fractional immunization in networks *Proc. 2013 SIAM Int. Conf. on Data Mining (SIAM)* pp 659–67
- [24] Lü L, Chen D, Ren X L, Zhang Q M, Zhang Y C and Zhou T 2016 *Phys. Rep.* **650** 1–63
- [25] Zhang H F, Wu Z X, Xu X K, Small M, Wang L and Wang B H 2013 *Phys. Rev. E* **88** 012813
- [26] Wang Z, Zhao D W, Wang L, Sun G Q and Jin Z 2015 *Europhys. Lett.* **112** 48002
- [27] Liu S, Perra N, Karsai M and Vespignani A 2014 *Phys. Rev. Lett.* **112** 118702
- [28] Morone F and Makse H A 2015 *Nature* **524** 65
- [29] Preciado V M, Zargham M, Enyioha C, Jadbabaie A and Pappas G J 2014 *IEEE Trans. Control Netw. Syst.* **1** 99–108
- [30] Preciado V M, Zargham M and Sun D 2014 A convex framework to control spreading processes in directed networks *48th Annual Conf. on Information Sciences and Systems (IEEE)* pp 1–6
- [31] Nowzari C, Preciado V M and Pappas G J 2016 Analysis and control of epidemics: a survey of spreading processes on complex networks *IEEE Control Syst. Mag.* **36** 26–46
- [32] Wang W, Liu Q H, Cai S M, Tang M, Braunstein L A and Stanley H E 2016 *Sci. Rep.* **6** 29259
- [33] Chen X, Wang R, Tang M, Cai S, Stanley H E and Braunstein L A 2018 *New J. Phys.* **20** 013007
- [34] Zhang H F, Xie J R, Tang M and Lai Y C 2014 *Chaos* **24** 043106
- [35] Zhao D, Wang L, Li S, Wang Z, Wang L and Gao B 2014 *PLoS One* **9** e112018
- [36] Ogura M and Preciado V M 2017 *Temporal Network Epidemiology* (Berlin: Springer) pp 241–66
- [37] Starnini M, Machens A, Cattuto C, Barrat A and Pastor-Satorras R 2013 *J. Theor. Biol.* **337** 89–100
- [38] Schneider C M, Mihaljev T, Havlin S and Herrmann H J 2011 *Phys. Rev. E* **84** 061911
- [39] Van Mieghem P, Stevanović D, Kuipers F, Li C, Van De Bovenkamp R, Liu D and Wang H 2011 *Phys. Rev. E* **84** 016101
- [40] Matamalas J T, Arenas A and Gómez S 2018 *Sci. Adv.* **4** eaau4212
- [41] de Arruda G F, Cozzo E, Peixoto T P, Rodrigues F A and Moreno Y 2017 *Phys. Rev. X* **7** 011014
- [42] Gómez S, Arenas A, Borge-Holthoefer J, Meloni S and Moreno Y 2010 *Europhys. Lett.* **89** 38009
- [43] Pan L, Wang W, Cai S and Zhou T 2019 *Phys. Rev. E* **100** 022316
- [44] Lee K M, Kim J Y, Cho W K, Goh K I and Kim I 2012 *New J. Phys.* **14** 033027
- [45] Cho W K, Goh K I and Kim I M 2010 (arXiv:1010.4971)
- [46] Freeman L C, Webster C M and Kirke D M 1998 *Soc. Netw.* **20** 109–18
- [47] Gleiser P M and Danon L 2003 *Adv. Complex Syst.* **6** 565–73

- [48] McAuley J and Leskovec J 2012 Learning to discover social circles in ego networks *Advances in Neural Information Processing Systems* pp 548–56
- [49] Coleman J, Katz E and Menzel H 1957 *Sociometry* **20** 253–70
- [50] Kunegis J 2013 KONECT: the Koblenz network collection *Proc. 22nd Int. Conf. World Wide Web* pp 1343–50