

# An Efficient Immunization Strategy for Community Networks

Kai Gong<sup>1,2</sup>, Ming Tang<sup>1,2,3\*</sup>, Pak Ming Hui<sup>2\*</sup>, Hai Feng Zhang<sup>4</sup>, Younghae Do<sup>3</sup>, Ying-Cheng Lai<sup>3,5</sup>

**1** Web Sciences Center, University of Electronic Science and Technology of China, Chengdu, People's Republic of China, **2** Department of Physics, The Chinese University of Hong Kong, Shatin, Hong Kong, People's Republic of China, **3** Department of Mathematics, Kyungpook National University, Daegu, South Korea, **4** School of Mathematical Science, Anhui University, Hefei, People's Republic of China, **5** School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona, United States of America

## Abstract

An efficient algorithm that can properly identify the targets to immunize or quarantine for preventing an epidemic in a population without knowing the global structural information is of obvious importance. Typically, a population is characterized by its community structure and the heterogeneity in the weak ties among nodes bridging over communities. We propose and study an effective algorithm that searches for bridge hubs, which are bridge nodes with a larger number of weak ties, as immunizing targets based on the idea of referencing to an expanding friendship circle as a self-avoiding walk proceeds. Applying the algorithm to simulated networks and empirical networks constructed from social network data of five US universities, we show that the algorithm is more effective than other existing local algorithms for a given immunization coverage, with a reduced final epidemic ratio, lower peak prevalence and fewer nodes that need to be visited before identifying the target nodes. The effectiveness stems from the breaking up of community networks by successful searches on target nodes with more weak ties. The effectiveness remains robust even when errors exist in the structure of the networks.

**Citation:** Gong K, Tang M, Hui PM, Zhang HF, Do Y, et al. (2013) An Efficient Immunization Strategy for Community Networks. PLoS ONE 8(12): e83489. doi:10.1371/journal.pone.0083489

**Editor:** Attila Szolnoki, Hungarian Academy of Sciences, Hungary

**Received:** August 11, 2013; **Accepted:** November 4, 2013; **Published:** December 20, 2013

**Copyright:** © 2013 Gong et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was partially supported by the National Natural Science Foundation of China (Grants Nos. 11105025, 11005001) and the Fundamental Research Funds for the Central Universities (Grant No. ZYGX2012J075). P. M. Hui acknowledges the support of a Direct Grant for Research from the Faculty of Science at the Chinese University of Hong Kong in 2013–14. Y. Do acknowledges Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A2010067). K. Gong acknowledges the support from the Program of Outstanding PhD Candidate in Academic Research by UESTC (Grant No. YBXSZC20131027). Y.-C. Lai was supported by AFOSR under Grant No. FA9550-10-1-0083. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: tangminghuang521@hotmail.com (MT); pmhui@phy.cuhk.edu.hk (PMH)

## Introduction

Epidemics could lead to a serious loss in life and have a huge impact on the economy, as we witnessed during the outbreaks of SARS (severe acute respiratory syndromes) in 2003 and H1N1 Influenza A virus in 2009 [1–3]. The problem of preventing an epidemic in time is of great importance and it has attracted much attention from researchers across different fields [4–12]. While immunization and quarantine are the two basic measures [13–15], finding an efficient way of identifying the targets to be immunized or quarantined so as to suppress an infection effectively remains a pressing issue [16].

The hubs, individuals with high centrality, in complex networks are commonly believed to be the most influential nodes as they could affect their many neighboring nodes [17]. Kitsak *et al.* pointed out that the nodes with a high  $k$ -shell are more influential spreaders in some real networks [18]. Undoubtedly, it is important to identify such influential nodes, but an effective method would require *global* information about the network structure. In these global strategies, the influential hubs can be identified by centrality measures, such as the degree centrality [19,20], eigenvector centrality [21], and betweenness centrality [22,23]. Such global strategies, however, become impractical for large-scale networks. Another type of method requires only *local* information. The

acquaintance immunization strategy (henceforth labeled as ACQ [24] in which random acquaintances of randomly chosen nodes are immunized is an example of local search algorithms. ACQ was shown to be an efficient algorithm for large-scale networks with broad-degree distributions [24].

*Community structure* at the mesoscale level is ubiquitous in a variety of real complex systems [25,26] such as Facebook [27,28] and Twitter [29], which plays an important role in the dynamics of epidemics [30–35]. In the presence of communities, the weak ties connecting a pair of nodes belonging to different communities, called the *bridge nodes* [36–38], provide the pathways for information and diseases to propagate from one community to another. These bridge nodes were found to be more important than the hubs in diffusing information through community networks [39–41]. Therefore, identifying the bridge nodes in community networks are crucial in preventing epidemic outbreaks [42–44]. Deterministic and stochastic algorithms have been designed to search for the bridge nodes [45]. Chen *et al.* [46] proposed an immunization strategy based on an equal graph partitioning algorithm to identify the minimum group that separates a network into several clusters of approximately equal size. The role of the minimum separator group is similar to that of the bridge nodes connecting different communities. The algorithm

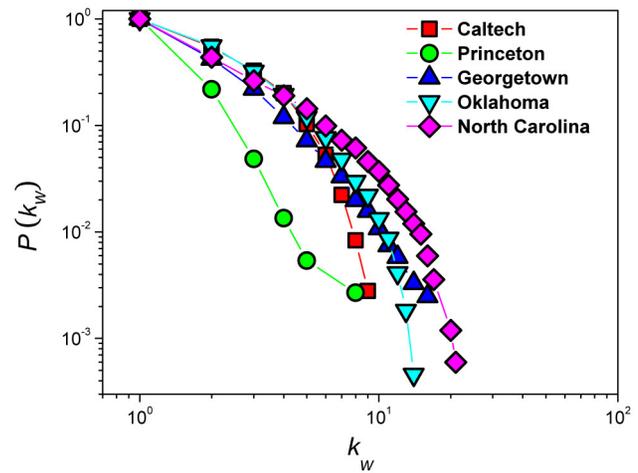
is effective in that only a small immunization ratio is needed, which is the fraction of immunized nodes. However, the algorithm requires the structural information of the whole network. Similar to other global strategies, it becomes impractical due to the high computing cost and, more fundamentally, the lack of the complete information in large networks. Salathe and Jones [45] proposed the *community bridge finder* (CBF), which is a stochastic algorithm to search for the bridge nodes based only on local structural information, and found that the method is more efficient than immunization strategies targeting different kinds of hubs.

Community networks typically exhibit a heterogeneous distribution in the number of weak ties originating from a bridge node [36,47]. In the worldwide air traffic network, for example, bridge nodes can be divided into four classes: nonhub connector nodes, nonhub kinless nodes, connector hubs, and kinless hubs [48]. It was found that immunizing bridge hubs that connect a community to many other communities could protect a network efficiently against epidemics [49]. Identifying the bridge hubs with a larger number of weak ties, however, is quite challenging as the mesoscopic-scale community structure is difficult to resolve for large-scale networks [50]. It is against this backdrop that we propose an immunization strategy that identifies the bridge hubs by random walks in an expanding friendship circle. The strategy, which we call bright-hub detector (BHD) relies only on local structural information. We compare results obtained by BHD with ACQ and CBF strategies and find that it performs better in simulated networks and empirical community networks generated from real data. Simulation results also show that our strategy has the merit of being robust against noise.

## Results

To illustrate the community structure in real-world networks and the necessity of an algorithm that focuses on identifying the bridge hubs, we have analyzed the heterogeneity of bridge nodes in the networks of students in five universities in the US using Facebook data. Details on constructing the networks from data are given in the Materials and Methods section. Figure 1 shows the cumulative probability density  $P(k_w)$  that a bridge node has a degree  $k_w$  or higher in the students' Facebook networks of Caltech, Princeton, Georgetown, Oklahoma, and North Carolina. The distributions indicate that it is important to identify the bridge nodes with more weak ties, in addition to identifying only the bridge nodes as in CBF.

We present results of three different strategies (ACQ, CBF and BHD) on choosing immunizing nodes in simulated and empirical community networks within the susceptible-infected-recovered (SIR) epidemiological model. Briefly, simulated networks of different modularity  $Q$  are generated by randomly connecting communities, which are themselves random networks, following the algorithm in Ref. [45]. Empirical networks are constructed from Facebook data of students in each of five US universities, with a link defined by two students being online friends and belonging to the same dormitory or the same major in the same year of study. Table 1 lists the properties of these empirical networks. Details about network construction, the three strategies, and SIR dynamics are given in the Materials and Methods section. The ACQ, CBF and BHD strategies rely only on local structural information. In ACQ [24], a node is randomly picked and then a neighbor of the chosen node is randomly selected for immunization. The algorithm has a higher chance to immunize the hubs. In CBF [45], a step forward in a self-avoiding walk is checked for a bridge between communities by examining the existence of links or paths from the neighbors of the last node back to the nodes



**Figure 1. Cumulative distributions of the number of weak ties in five empirical networks.** For every network, the community structure is detected by the method proposed by Newman and Girvan [55]. Weak ties and bridge nodes are then identified. The number of weak ties emanating from each bridge node is recorded to give the cumulative distribution. Results are shown for students' networks for Caltech (red squares), Princeton (green circles), Georgetown (blue up triangles), Oklahoma (cyan down triangles) and North Carolina (magenta diamonds).  
doi:10.1371/journal.pone.0083489.g001

visited in the trail of the self-avoiding walk. The node before crossing the bridge is then identified as the bridge node and immunized. In addition, the supplementary rule of immunizing a node that has been visited twice by self-avoiding walks would pick up some hubs. Here, we implemented CBF as given in Ref. [45]. The bridge-hub detector (BHD) that we proposed in this paper *extends* the self-avoiding searching scheme to examining the overlap and the existence of links from all the neighbors of the last node back to the *union of the friendship circles* of all the nodes in the trail of the walk. A pair of nodes, a bridge node and a bridge hub, are searched for immunization via a self-avoiding walk (Readers are referred to the Materials and Methods section for more information). In the SIR model, the parameters  $\beta$  and  $\mu$  are the transmission rate and the recovery probability, respectively. The extent of an epidemic is characterized by the final epidemic ratio  $R$ , the fraction of the population ever infected at the end of the epidemic, and the peak prevalence  $I_m$ , the highest density of infected nodes in the population at a time during the epidemic.

To compare the performances of ACQ and CBF with that of BHD, we have carried out the SIR dynamics on simulated networks with different modularity. For a given immunization coverage  $f$ , i.e., fraction of immunized nodes, the final epidemic ratio  $R$  is obtained for each of the three strategies. The differences  $R_{ACQ} - R_{BHD}$  and  $R_{CBF} - R_{BHD}$  would indicate the performance of ACQ and CBF relative to BHD. Figure 2 shows the differences for various combinations of the modularity  $Q$  and coverage  $f$ . The differences are positive over most of the  $Q$ - $f$  space, with BHD leading to about 6.1% (2.8%) of fewer nodes being infected than ACQ (CBF). The results indicate that BHD is, in general, more effective in suppressing an epidemic. There are patches in the  $Q$ - $f$  space that BHD is out-performed by ACQ and CBF. This is particularly apparent for networks with a strong community structure characterized by a high modularity. For example, both ACQ and CBF perform better in networks with  $Q \in [0.95, 0.97]$ . In these cases, an epidemic outbreak is mostly restricted in a local community and thus the better strategies are those that could

**Table 1.** Structural Properties of the five empirical networks.

Network	N	E	M	Q	r	H	C	$\langle d \rangle$	$\langle k \rangle$
Caltech	647	7,047	13	0.675	0.187	6.517	0.443	3.166	21.784
Princeton	5,580	156,935	8	0.817	0.039	6.366	0.298	4.316	56.252
Georgetown	8,183	245,510	42	0.814	0.106	6.280	0.268	4.218	60.005
Oklahoma	13,515	276,126	67	0.774	0.327	8.240	0.266	4.108	40.862
North Carolina	15,425	332,314	70	0.779	0.217	6.941	0.233	4.346	43.086

Structural properties including the network size (N), number of edges (E), the community number (M), modularity (Q) [55], degree assortativity (r) [56], degree heterogeneity ( $H = \langle k^2 \rangle / \langle k \rangle^2$ ), clustering coefficient (C), average shortest path length ( $\langle d \rangle$ ) and average degree ( $\langle k \rangle$ ) are tabulated for each of the five empirical networks.

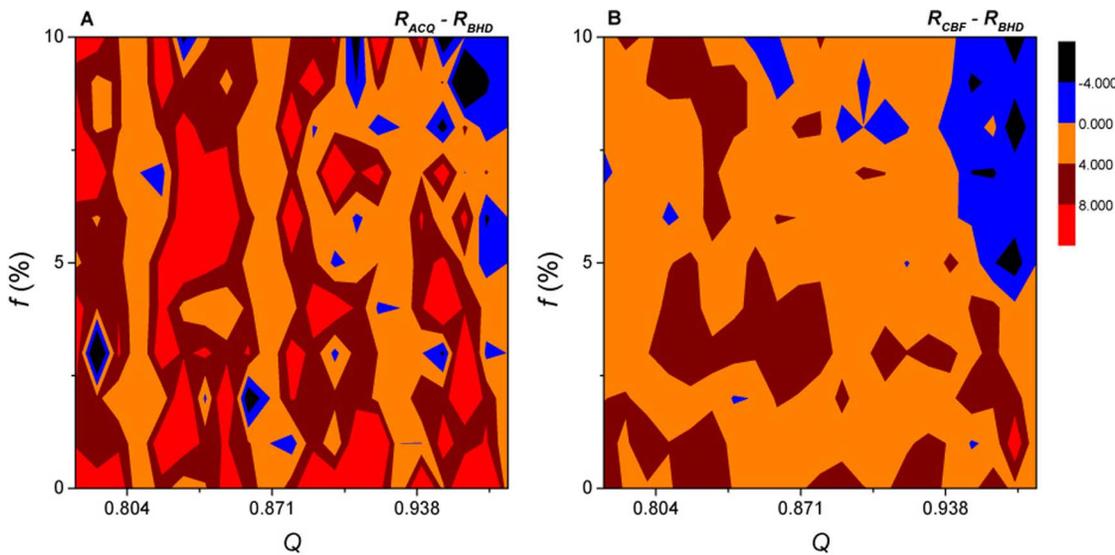
doi:10.1371/journal.pone.0083489.t001

confine and suppress an epidemic to within the community that the outbreak starts. Immunizing the hubs with large degrees as in ACQ and CBF can prevent epidemic spreading within a community more efficiently. BHD spends some of the immunization coverage on removing nodes that belong to neighboring communities rather than the community of the outbreak itself and this becomes less effective in saving the nodes within the outbreak community from infection than ACQ and CBF.

Figure 3 shows the results of  $R_{ACQ} - R_{BHD}$  and  $R_{CBF} - R_{BHD}$  obtained for the empirical networks. In most cases, the final epidemic ratio is smaller for BHD than ACQ and CBF. For example,  $R_{BHD}$  is on average 8.5% (6.0%) smaller than  $R_{ACQ}$  ( $R_{CBF}$ ) for the Caltech network, 4.3% (2.3%) smaller for the Princeton network, and 3.2% (2.2%) smaller for the Georgetown University network. For the Oklahoma University network,  $R_{BHD}$  is about 3.0% (1.4%) smaller than  $R_{ACQ}$  ( $R_{CBF}$ ) when the immunization coverage is not too small, e.g.  $f = 20\%$  ( $f = 22\%$ ). Similarly,  $R_{BHD}$  is about 2.9% (1.0%) smaller than  $R_{ACQ}$  ( $R_{CBF}$ ) at  $f = 20\%$  ( $f = 26\%$ ) for the North Carolina University network. In the Oklahoma and North Carolina networks, BHD is out-

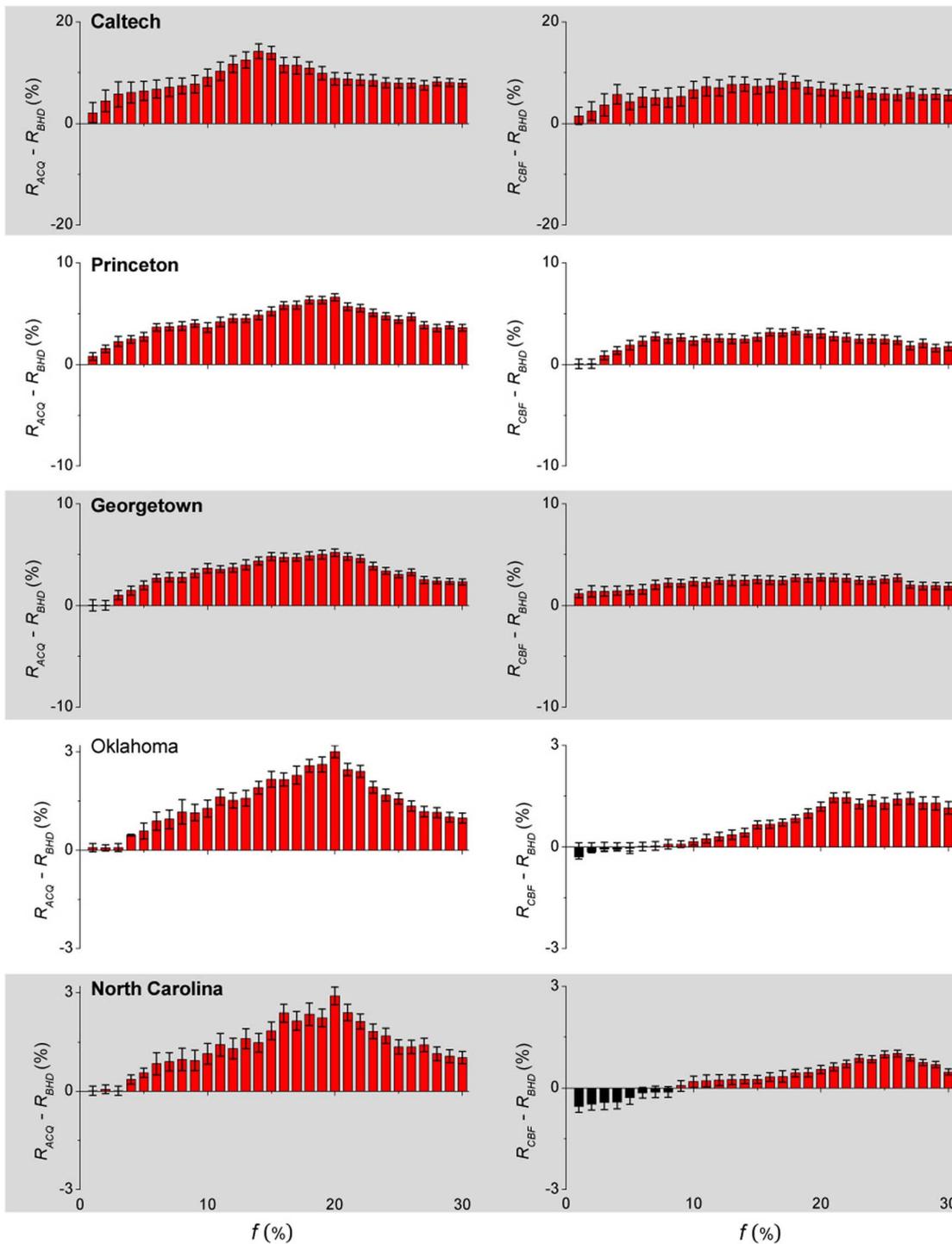
performed by CBF when the immunization coverage is low. The ability of CBF to immunize the bridge nodes and some hubs makes it more effective in networks with a high assortativity and heterogeneity. As shown in Table 1, the degree assortativity  $r$  and heterogeneity  $H$  are  $\approx 0.327$  ( $\approx 0.217$ ) and  $\approx 8.240$  ( $\approx 6.941$ ) for the Oklahoma (North Carolina) University network. For an immunization coverage of  $f \approx 20\%$ , BHD performs better in all empirical networks than both ACQ and CBF, with a maximum in  $R_{ACQ} - R_{BHD}$  near this coverage. When the coverage is higher (e.g.  $f > 20\%$ ), the differences in  $R$  for the three algorithms diminish, as the many immunized nodes would tend to be effective in breaking up a community and suppress an epidemic regardless of the algorithm. The results for the differences in peak prevalence are given in Figure S1 and the features are essentially the same as those in Figure 3.

Whether an algorithm is effective in preventing an epidemic is closely related to its effectiveness in breaking up the network after immunizing or removing a certain fraction of nodes [51]. Let  $G_\alpha(f)$ , with  $\alpha$  denoting ACQ, CBF and BHD, be the size of the giant component of the resulting network using the algorithm  $\alpha$  and an immunization coverage  $f$ , and  $G(0)$  be the size of the giant



**Figure 2. Comparison of efficacy of immunization algorithms in simulated networks.** The difference in the final epidemic ratios (a) (left panel)  $R_{ACQ} - R_{BHD}$  between ACQ and BHD, and (b) (right panel)  $R_{CBF} - R_{BHD}$  between CBF and BHD, are shown for simulated networks with different network modularity  $Q$  and immunization coverage  $f$ . The colors indicate the differences in percentages (see color codes). Results are obtained by averaging over  $2 \times 10^3$  realizations for each pair of  $Q$  and  $f$  values. The parameters associated with the SIR dynamics are  $\beta = 0.08$  and  $\mu = 0.2$ .

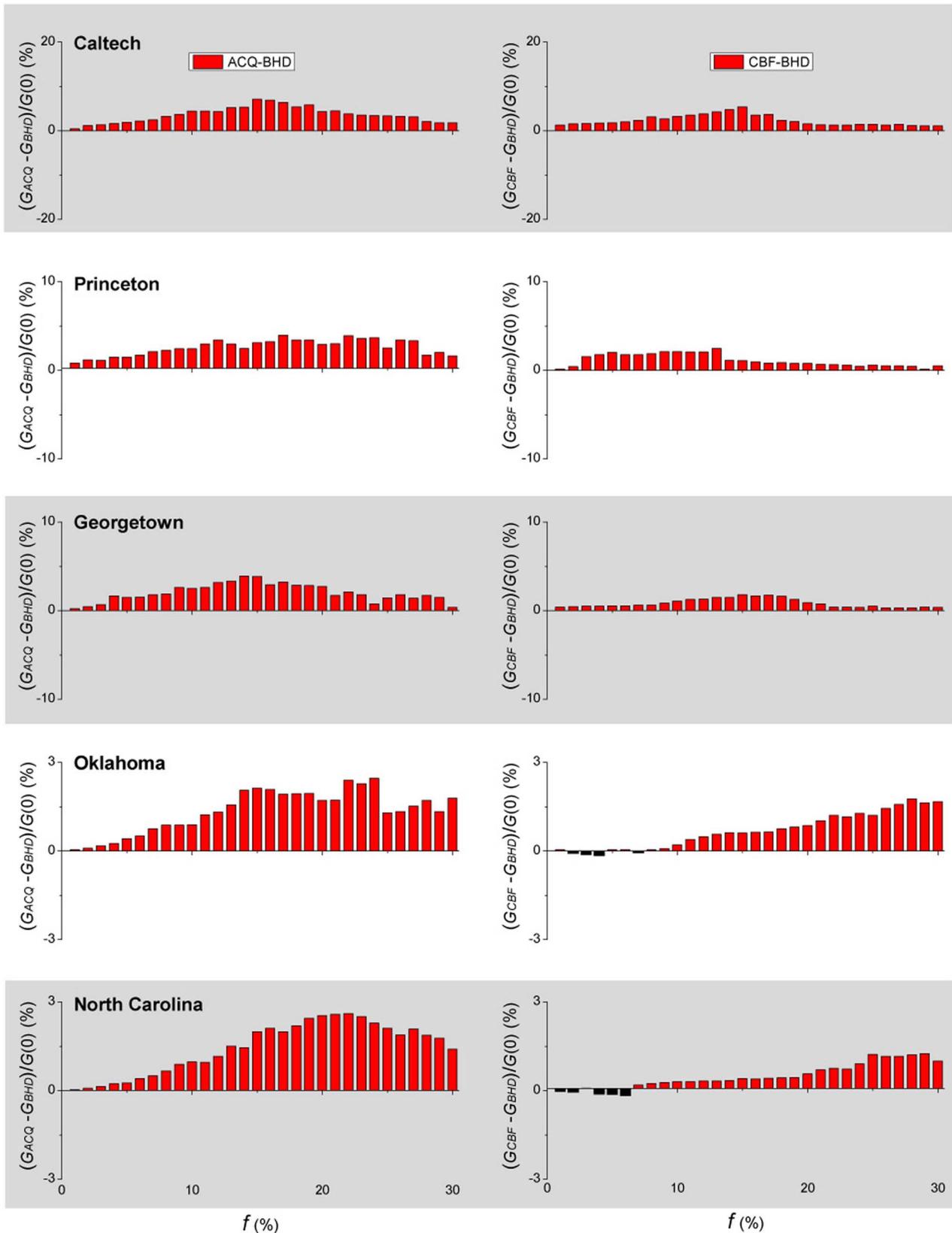
doi:10.1371/journal.pone.0083489.g002



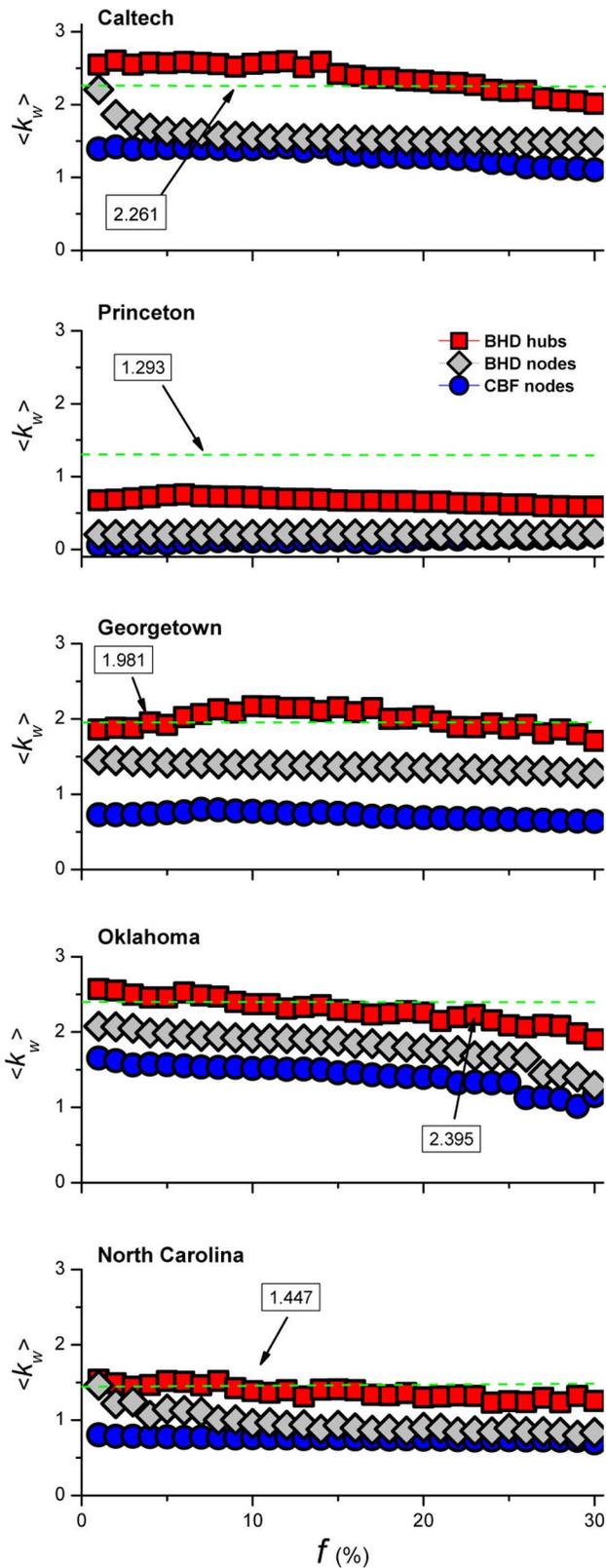
**Figure 3. Comparison of efficacy of immunization algorithms in empirical networks.** The differences in the final epidemic ratios  $R_{ACQ} - R_{BHD}$  (left panel) and  $R_{CBF} - R_{BHD}$  (right panel) using different immunization algorithms are shown for each of the five empirical networks as a function of the immunization coverage  $f$ . A positive value indicates that BHD is more effective than the other algorithms. Results are obtained by averaging over  $10^4$  realizations for each value of  $f$ . doi:10.1371/journal.pone.0083489.g003

component of the network before applying an immunization algorithm. The relative difference  $\Delta G(f)/G(0) = [G_{ACQ/CBF}(f) - G_{BHD}(f)]/G(0)$  is a measure of how effective BHD breaks up a network relative to ACQ (and CBF). Figure 4 shows the results of  $\Delta G(f)/G(0)$  for the five empirical networks. The positive

$\Delta G(f)/G(0)$  values in almost all the cases indicate that BHD is more effective. The results show a similar trend as a function of  $f$  as in the differences in final epidemic ratio shown in Figure 3. Percentage wise,  $\Delta G(f)/G(0)$  is smaller than  $R_{ACQ} - R_{BHD}$  and  $R_{CBF} - R_{BHD}$  for the same values of  $f$  in the same empirical



**Figure 4. Comparison of giant components of different immunization algorithms in empirical networks.** The difference in the sizes of the giant component  $(G_{ACQ} - G_{BHD})/G(0)$  (left panel) and  $(G_{CBF} - G_{BHD})/G(0)$  (right panel) using different immunization algorithms are shown for each of the five empirical networks as a function of the immunization coverage  $f$ , where  $G(0)$  is the size of the giant component before an immunization algorithm is applied. A positive value indicates that BHD is more effective in breaking up the network. Results are obtained by averaging over  $10^4$  realizations for each value of  $f$ .  
doi:10.1371/journal.pone.0083489.g004



**Figure 5. Comparison of average number of weak ties among immunized bridge hubs identified by BHD, bridge nodes identified by BHD and bridge nodes identified by CBF. BHD identifies a pair of nodes for immunization via a self-avoiding walk algorithm. One node is a bridge hub and another a bridge node.**

The average number of weak ties among these two types of immunized nodes [red squares (BHD hubs) and gray diamonds (BHD nodes)] are shown for different values of immunization coverage  $f$ , together with the results from immunized bridge nodes identified by CBF (blue circles). Results are obtained by averaging over  $10^4$  realizations for each value of  $f$ . For comparison, the results based on the method proposed by Newman and Girvan [55] are shown as the green dashed lines. doi:10.1371/journal.pone.0083489.g005

network. It is a result of the removal of some bridge hubs (bridge nodes with more weak links to other communities), by BHD, which would result in a network that has a smaller mean degree. This is indeed the case, as shown in Figure S2. The geometry of the resulting network also affects the duration of an epidemic. In particular, resulting networks with smaller mean degree lead to a longer duration, as the disease can only spread by infecting one node after another. BHD is therefore expected to have a longer epidemic duration, despite a smaller final epidemic ratio and prevalence. Figure S3 shows that the epidemic duration is slightly longer for BHD, as compared with those with ACQ and CBF.

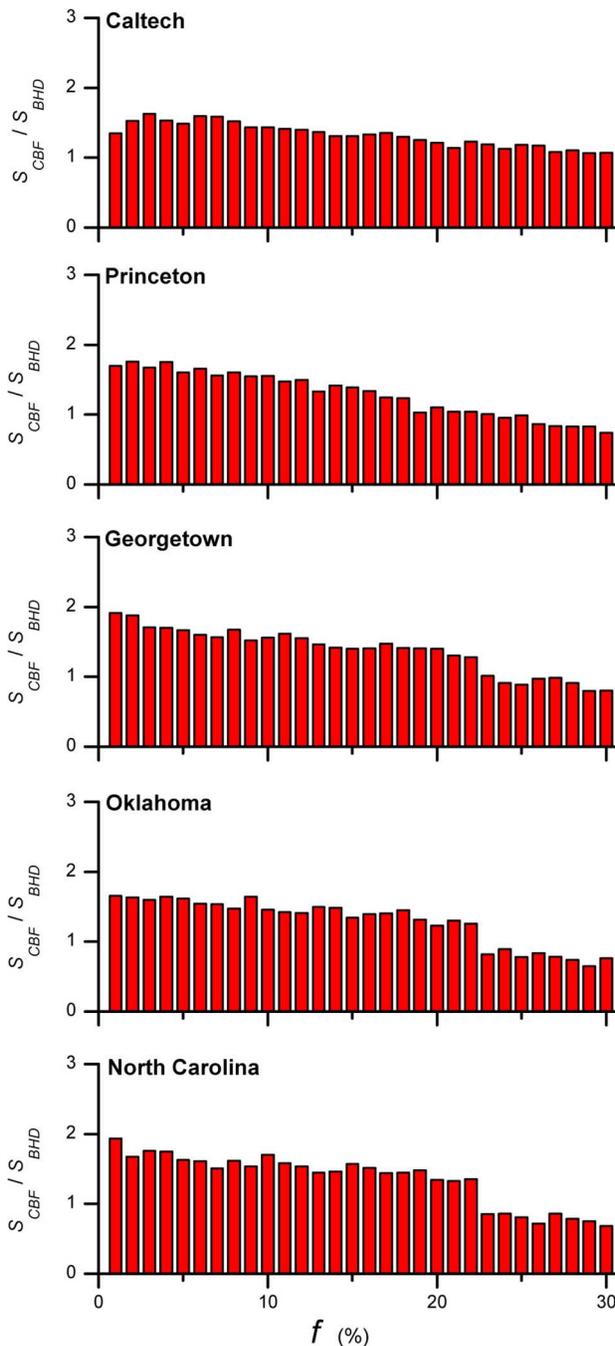
The effectiveness of BHD is further illustrated in Figure 5, which compares the number of weak ties of the bridge hubs and bridge nodes identified for immunization using BHD with that of the bridge nodes identified by CBF. The results show that BHD indeed can identify the nodes with more weak links for removal. In addition, the bridge nodes identified by BHD carry more weak ties than those identified by CBF, further improving the effectiveness. In the Princeton University network, however, the large mean community size (see Table 1) makes it difficult for both BHD and CBF to search for the bridge nodes.

A more efficient algorithm is one that visits fewer nodes before identifying the targeted nodes for immunization. Let  $S_{CBF}$  and  $S_{BHD}$  be the numbers of nodes visited by the self-avoiding walks before getting at the targeted nodes. Figure 6 gives  $S_{CBF}/S_{BHD}$  for the five empirical networks. For small immunization coverage ( $f \leq 20\%$ ),  $S_{CBF}/S_{BHD} \approx 1.5$ , indicating that BHD is a more efficient algorithm.

Another aspect of an effective algorithm is the robustness to errors or noise in network information. Such errors are common in social networks due to, for example, the inconsistency for two individuals to express their relationship [52]. Using an empirical network, we have tested the robustness of BHD by adding or removing a number of links from the network. SIR dynamics is then studied on the modified network without and with the BHD algorithm. The resulting final epidemic ratios are denoted by  $R_0$  and  $R_{BHD}$ , respectively. Figure 7 shows the results of  $R_{BHD} = (R_0 - R_{BHD})/R_0$  for different numbers of links added or removed, with a constant immunization coverage of 5%. The results show that BHD performs well even when the networks are modified randomly. More results corresponding to other values of immunization coverage are shown in Figure S4.

**Summary and Discussion**

The heterogeneous distribution of weak ties in real-world community networks points to the importance of the bridge hubs in the control of transmission of information or diseases but these nodes are difficult to identify. We have proposed and studied the effectiveness of a bridge-hub detector strategy. It is a local strategy searching for bridge nodes and bridge hubs based on the idea of referencing to an expanding friendship circle as a self-avoiding walk proceeds. We applied BHD to simulated networks and empirical networks among students in five US universities constructed by using social network data. In general, BHD is more effective in preventing an epidemic when compared with



**Figure 6. Comparison of number of nodes visited before identifying immunization nodes in CBF and BHD.** The ratio  $S_{CBF}/S_{BHD}$  is shown of the number of nodes visited by the self-avoiding walks in CBF and BHD for achieving an immunization coverage  $f$ . A large  $S$  implies a longer search for immunization nodes. The results show that BHD identifies the immunization nodes faster than CBF for  $f \leq 20\%$ . Results are obtained by averaging over  $10^4$  realizations for each value of  $f$ . doi:10.1371/journal.pone.0083489.g006

other local immunization strategies such as ACQ and CBF, for a practical range of immunization coverage. It gives a smaller final epidemic ratio and peak prevalence. Its effectiveness can be attributed to better identification of immunized nodes that carry more weak ties than those picked up by ACQ and CBF. BHD thus

breaks up the community networks more effectively and suppresses an epidemic. Although BHD and CBF are both based on self-avoiding walks, BHD is more efficient in that it identifies the immunization nodes after visiting fewer nodes. The good performance of BHD remains robust even when errors exist in the structure of the underlying network.

In general, BHD is capable of identifying bridge nodes with more weak ties and it is particularly useful in dealing with community networks with a broad distribution in the number of weak ties among the bridge nodes. It can be readily generalized to other tasks. For example, mistakenly chosen nodes could be reduced by requiring that a node must have been identified twice or more times before it is immunized. The algorithm can also be applied to overlapping [50] and time-varying [53] community networks.

## Materials and Methods

We have studied the effectiveness of our local immunization strategy in simulated and empirical community networks using the susceptible-infected-recovered (SIR) epidemiological model. Here, we give the details with respect to the following issues: community-network construction, bridge-hub detector algorithm, and SIR dynamics.

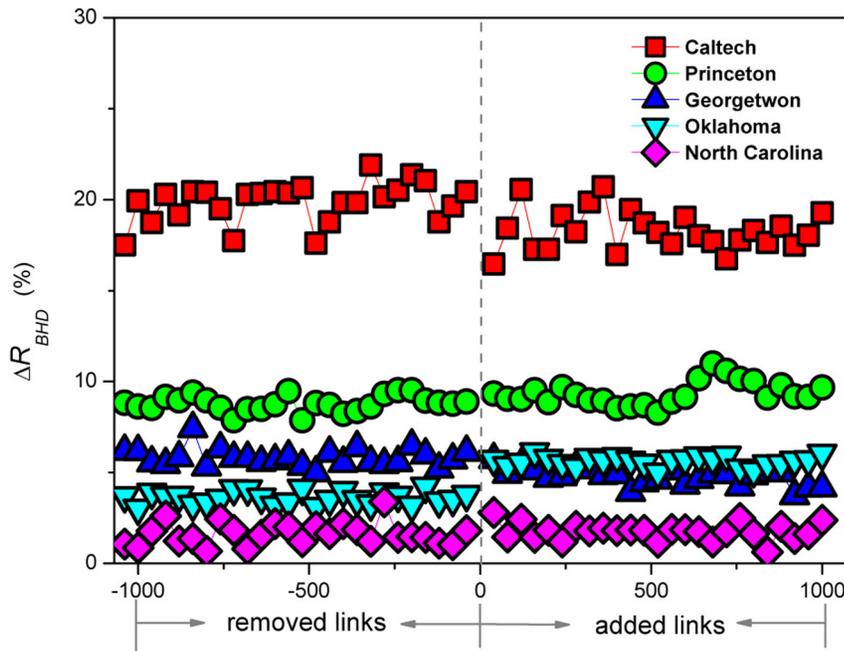
### Simulated and empirical community networks

The simulated community networks of different modularity  $Q$  are generated by the algorithm given in Ref. [45]. There are  $m$  independent random communities. In each community,  $n$  nodes are randomly connected so that the mean degree is  $\langle k_{in} \rangle$ . These communities are then connected randomly by  $E_{out}$  links. The simulated community network thus has a total of  $mn$  nodes and  $mn\langle k_{in} \rangle + E_{out}$  undirected links. We used the same set of parameters as in Ref. [45], namely  $m=50$ ,  $n=40$ ,  $\langle k_{in} \rangle=8$ , and  $E_{out}=2000$ . After generating the network, the modularity  $Q$  can be evaluated according to the definition given in Ref. [25]. The modularity  $Q$  can be varied by rewiring some inter-community links into intra-community links, following the rewiring procedures given in Ref. [45]. While the modularity increases with rewiring, the degree heterogeneity remains nearly unchanged.

Empirical community networks are constructed using the collegiate social network data from Facebook (www.facebook.com) studied in Ref. [27]. The data of five universities in the US are studied (Available: <https://code.google.com/p/socialnetworksimulation/>). They include data of students from the California Institute of Technology (Caltech), Princeton University, Georgetown University, University of Oklahoma, and University of North Carolina. The data provide information on the dormitory, majoring subject and year of class of the members. Based on the data, a community network is constructed [45] by linking up two members when (i) they are online friends of each other, and (ii) they belong to the same dormitory or the same major in the same year of study. The largest connected component of the network is then retained for carrying out our study. Basic statistical properties of the five empirical networks are given in Table 1. These networks exhibit the small-world character with a high clustering coefficient and a short average path length. The high modularity indicates a strong community feature.

### Local Immunization Strategies and Bridge-Hub Detector

The present paper compares results based on our bridge-hub detector algorithm (BHD) with those obtained by other local strategies based on acquaintance immunization (ACQ) and the



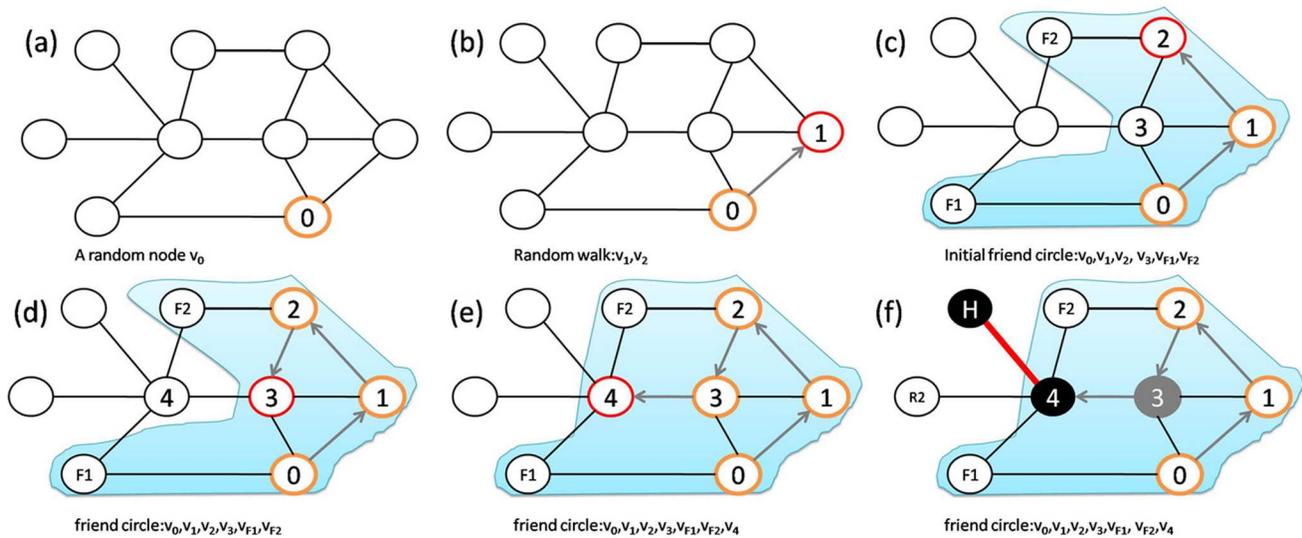
**Figure 7. Robustness of BHD in networks with noise as modeled by random addition and removal of links.** The quantity  $\Delta R_{BHD} = (R_0 - R_{BHD})/R_0$  is shown as a function of the number of links randomly added to or removed from an empirical network, where  $R_{BHD}$  and  $R_0$  are the final epidemic ratios when BHD is applied and not applied to the modified network, respectively. The robustness of BHD is indicated by the relatively stable values of  $\Delta R_{BHD}$ . Results are shown for each of the five empirical networks as labeled. The immunization coverage is  $f = 5\%$ . The SIR parameters are  $\beta = 0.08$  and  $\mu = 0.2$ . Results are obtained by averaging over  $10^4$  realizations for each value of added and removed links. doi:10.1371/journal.pone.0083489.g007

community bridge finder (CBF). Before discussing the BHD algorithm, we outline the ideas behind ACQ and CBF.

The ACQ algorithm immunizes randomly chosen acquaintances of randomly chosen nodes [24]. In ACQ, a node is picked randomly and then a neighbor of the chosen node is picked randomly for immunization. The procedure is then repeated until the desired immunization coverage is achieved. Therefore, a node that is a neighbor of many other nodes, i.e., a hub, will have a higher chance to be chosen for immunization. The algorithm does not require any information about the degree of the nodes. It is efficient for large networks with a broad degree distribution. A similar algorithm that requires a node to be chosen two times according to the ACQ random-pick processes (ACQ2) before it is immunized has also been studied [24]. Comparing with ACQ, ACQ2 has a higher probability of immunizing the hubs.

Bridge nodes are important for transmission through community networks, but ACQ does not aim at the bridge nodes. In contrast, CBF is designed to search for bridge nodes [45]. It is an algorithm based on self-avoiding walks. To identify a bridge and its associated bridge node for immunization, a self-avoiding walk is executed starting from a randomly chosen node  $v_0$ . The following procedure is taken after every step when the walk has visited three or more nodes (see Figure 8). For a walk after  $t$  steps ( $t \geq 2$ ), the set of all nodes visited in the  $t$  steps so far is denoted by  $\{v_{t'}\}$  for all  $t' = 0, 1, \dots, t$ . Thus,  $v_t$  is the node that the walk currently locates after step  $t$  and  $v_{t-1}$  is the node visited at step  $t-1$ . The first check is to examine whether the node  $v_t$  has a link or several links to the nodes in the set  $\{v_{t'}\}$  other than the link between  $v_t$  and  $v_{t-1}$ . If it is the case, the self-avoiding walk continues to step  $t+1$ . If not, then  $v_{t-1}$  is considered a possible target of a bridge node. To determine whether the node  $v_{t-1}$  is a bridge node, two nodes are randomly chosen among all the possible nodes that the walk could go in step  $t+1$ , i.e., two neighbors of the node  $v_t$  are randomly

chosen from all the neighbors (except the node  $v_{t-1}$  due to the self-avoiding restriction of the walk). In the case that there is only one neighbor to choose from, the only neighbor will be considered. If there exists a path from any of the two chosen nodes back to any node in the set  $\{v_{t'}\}$ , then the node  $v_{t-1}$  will not be regarded as a bridge node and the walk moves on from  $v_t$  to some node  $v_{t+1}$ . If there exists no path back to the set  $\{v_{t'}\}$  from both of the chosen nodes,  $v_{t-1}$  is regarded as a bridge node that connects two communities and it will be immunized. A new self-avoiding walk will start again at another randomly chosen node and the procedure is repeated until the desired immunization ratio is achieved. The idea behind CBF is that a community is formed by a circle of close friends, and thus when the two randomly chosen neighbors of  $v_t$  cannot be traced back to the community that  $v_{t-1}$  belongs to, the link between  $v_{t-1}$  and  $v_t$  is likely to be a bridge between two communities and so the node  $v_{t-1}$  is a bridge node. In practice, two additional checks are implemented to shorten the computing time [45]. Firstly, the number of nodes registered in a running path is kept at the length of ten, using the latest ten nodes visited. Secondly, the number of visits by any random walk for each node is recorded. When the number  $k$  of visits equals a certain number ( $k=2$  in Ref. [45]), the node is immunized. We have implemented CBF following the algorithm in Ref. [45], where it has been shown that, without prior knowledge of the community structure, CBF is more efficient than ACQ and other immunization strategies that target at the different kinds of hubs. In its design, CBF does not aim at identifying the bridge hubs. In real-world community networks such as those in Facebook [47], worldwide air traffic network [48], and the five collegiate community networks studied here, the heterogeneous distribution in the number of weak ties among bridge nodes indicates the importance of bridge hubs. Figure 9 shows an example of a bridge hub through a visualization of the students' network in Caltech.



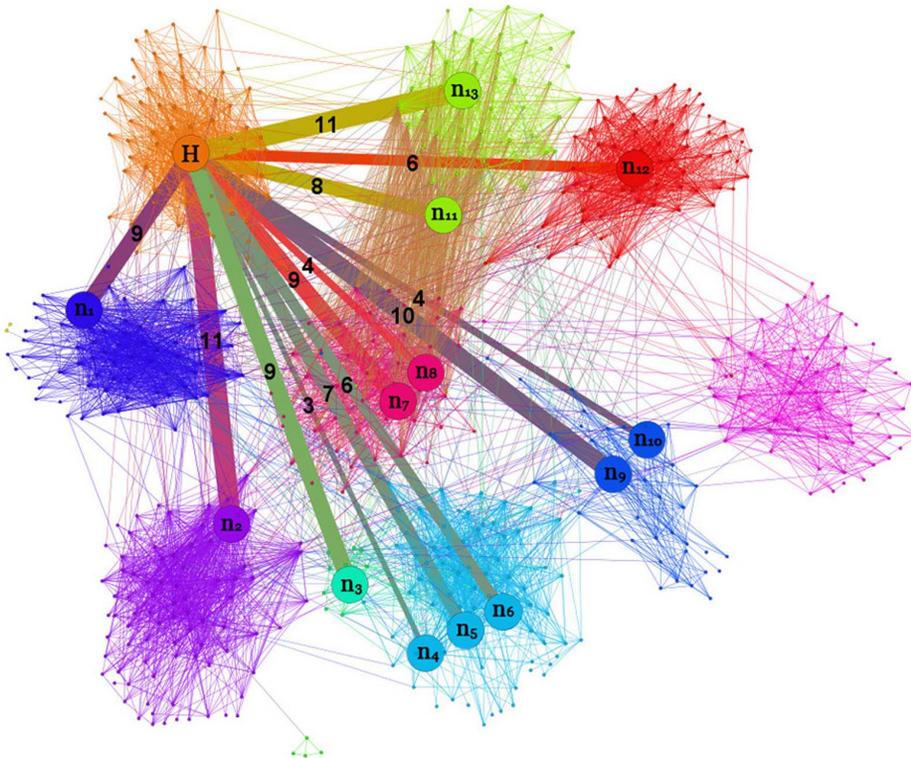
**Figure 8. Schematic illustration of CBF strategy and BHD strategy. The CBF strategy:** (a) A self-avoiding walk starts at a randomly chosen node  $v_0$ . (b)-(c) The walk visited  $\{v_0, v_1, v_2\}$  after two steps. The node  $v_2$  does not have any other links back to  $v_0$  and  $v_1$  other than the link that took the walk from  $v_1$  to  $v_2$ . The node  $v_1$  is a potential candidate of a bridge node. Two neighbors of  $v_2$ , namely  $v_3$  and  $v_{F2}$ , are randomly picked and each is examined for connections to the visited set of nodes. As  $v_3$  has links with  $v_0$  and  $v_1$ , the target node  $v_1$  is dismissed as a bridge node and the walk continues. (d) The walk moves to  $v_3$  after three steps. The node  $v_3$  has links to previously visited nodes  $v_0$  and  $v_1$ , and thus the walk continues. (e) The walk moves to  $v_4$  after four steps. The node  $v_4$  does not have any other links back to previously visited nodes other than the link that took the walk from  $v_3$  to  $v_4$ . The node  $v_3$  is a potential target of a bridge node. Two neighbors of  $v_4$  are randomly chosen. (f) If  $v_{R2}$  and  $v_H$  are chosen, these nodes do not connect back to the previously visited nodes and  $v_3$  is identified as a bridge node and immunized. **The BHD strategy:** (a) A self-avoiding walk starts at a randomly chosen node  $v_0$ . (b)-(c) The walk visited  $v_0, v_1$  and  $v_2$  after two steps. The set  $F_2$  is the union of all the neighboring nodes or friendship circles of  $v_0, v_1$  and  $v_2$ , as shaded in (c). (d) The walk moves to  $v_3$  after three steps. The friendship circle  $f_3$  of node  $v_3$  consists of  $v_0, v_1, v_2$ , and  $v_4$ . As all the nodes in  $f_3$  either belong to  $F_2$  or have a link to at least a node in  $F_2$ , the node  $v_3$  is not a potential target for immunization. (e) The union of friendship circles is updated to  $F_3 = f_3 \cup F_2$  as shaded. The walk continues and reaches node  $v_4$  after four steps. The friendship circle  $f_4$  of  $v_4$  consists of  $v_3, v_{F2}, v_{F1}, v_{R2}$  and  $v_H$ , among them  $v_{R2}$  and  $v_H$  do not belong to  $F_3$  and do not have a link to nodes in  $F_3$ . The node  $v_4$  is then identified as a bridge node for immunization. In addition, among those nodes in  $f_4$  that cannot be linked back to  $F_3$ , one node, e.g.,  $v_H$ , is randomly chosen and identified as a bridge hub for immunization. doi:10.1371/journal.pone.0083489.g008

We propose a bridge-hub detector (BHD) for more effective immunization in community networks. It is a local algorithm based on an expanding friendship circle and it does not rely on prior knowledge of the community structure. As in CBF, the algorithm starts with a self-avoiding walk at a randomly chosen node. The following procedure is taken after every step when the walk has visited three or more sites (see Figure 8). Let  $v_t$  be the node that the walker locates after step  $t$  and  $f_t$  be the set of all neighbors of the node  $v_t$ . For a walk up to  $t$  steps ( $t \geq 2$ ), the following steps are carried out between the set  $f_t$  and the set  $F_{t-1} \equiv f_0 \cup f_1 \cup f_2 \cup \dots \cup f_{t-1}$ . The set  $F_{t-1}$  is the union of all the friendship circles of all the nodes visited by the walker up to the step  $t-1$  and it expands as the self-avoiding walk proceeds. Apparently, we have  $v_{t-1} \in f_t$ . If every other node in  $f_t$  either already belongs to the set  $F_{t-1}$  or has at least a link to the nodes within  $F_{t-1}$ ,  $v_t$  will not be taken as the target of immunization and  $F_t$  will be updated to  $F_t = f_t \cup F_{t-1}$ . The self-avoiding walk then moves on from  $v_t$ . Otherwise, there must be at least a node in  $f_t$  that is not a member in  $F_{t-1}$  and is not connected to the nodes in  $F_{t-1}$ . The node  $v_t$  is then targeted for immunization. In addition, among the nodes in  $f_t$  that do not belong to  $F_{t-1}$  and cannot be linked back to  $F_{t-1}$ , one node (call it  $v_H$ ) is randomly chosen for immunization. A pair of nodes,  $v_t$  and  $v_H$ , are immunized and the self-avoiding walk stops. A new walk is initiated at another randomly chosen node again. The procedure is repeated until the desired immunization ratio is achieved. Similar to the idea in ACQ, the immunized node  $v_H$  in every self-avoiding walk is likely to be a bridge hub in the presence of heterogeneous distribution of

weak ties among the bridge nodes, while  $v_t$  is a bridge node. The algorithm is illustrated schematically in Figure 8. Practically, we terminate a walk when it fails to identify immunization nodes after 20 steps. As in CBF, BHD is effective when the search for immunization nodes begins after  $t \geq 2$  steps. Figure S5 in Supporting Information gives results when the search begins after different numbers of steps. Through computational complexity analysis (see Text S1 in Supporting Information for more details), the algorithms ACQ, CBF and BHD take on the worst-case run times that go as  $O(fN)$ ,  $O(fN^3)$  and  $O(f\langle k \rangle N^3)$ , respectively. For the different algorithms working on systems with the same parameters (the same network size  $N$  and immunization ratio  $f$  etc.), ACQ is the fastest, followed by CBF and BHD is the slowest. On a sparse network with small  $\langle k \rangle$ , BHD will have almost the same computational complexity as CBF.

**Epidemic Dynamics**

The susceptible-infected-recovered (SIR) epidemiological model [54] is used to test the performance of different immunization strategies on community networks. In the model, each node in the network represents an individual that can be in one of three states: susceptible (S), infected (I), or refractory/recovered (R); and each link represents a connection that could spread a disease. When an immunization strategy is executed, a percentage of nodes are first removed from the network by implementing ACQ, CBF or BHD. The remaining nodes are set to the susceptible state. To initiate an infection, a node is randomly chosen and turned into the I state. In a time step, a susceptible node will be infected with a probability



**Figure 9. Visualizing the community structure and a bridge hub in the Caltech network.** The Caltech network can be divided into 13 communities based on the method of Newman and Girvan [55], as illustrated by the different colors. Also illustrated is a bridge hub  $H$  that carries the largest number of weak ties and connects with other bridge nodes labeled as  $n_1$  to  $n_{13}$ . The number on a weak tie shows the difference in the numbers of weak ties originated from the node  $H$  and the bridge node at the other end of the weak tie.  
doi:10.1371/journal.pone.0083489.g009

$1 - e^{(-\beta^* i)}$ , where  $i$  is the number of infected nodes among the connected neighbors of the node. Here,  $\beta$  is a parameter characterizing the transmission rate. Meanwhile, an infected node has a recovery probability  $\mu$  to turn into the R state. Once an individual is in the R state, there will be no further change. In the simulations, the parameters are taken to be  $\beta = 0.08$  and  $\mu = 0.2$  to ensure an outbreak and the state of every node is updated synchronously in every time step. The dynamics ends when all infected nodes have recovered. The resulting population contains nodes only in the S state and R state. We record the *final epidemic ratio*  $R$  corresponding to the density of R nodes at the end of the epidemic, the *peak prevalence*  $I_m$  corresponding to the highest density of infected nodes in the network during the epidemic, and the *duration of epidemic* before the dynamics ends, and compare these quantities for ACQ, CBF and BHD.

## Supporting Information

**Figure S1 Comparison of peak prevalence of immunization algorithms in empirical networks.** The differences in the peak prevalence (left panel)  $\Delta I_m = I_{m,ACQ} - I_{m,BHD}$  between ACQ and BHD, and (right panel)  $\Delta I_m = I_{m,CBF} - I_{m,BHD}$  between CBF and BHD are shown for each of the five empirical networks as a function of the immunization coverage  $f$ . A positive value indicates that BHD is more effective than the other algorithms. The comparison is similar in features with those shown in Figure 3 for the differences in the final epidemic ratios. Results are obtained by averaging over  $10^4$  realizations for each value of  $f$ . (EPS)

**Figure S2 Comparison of residual mean degrees of different immunization algorithms in empirical networks.** The differences in the mean degrees  $(\langle k \rangle_{ACQ} - \langle k \rangle_{BHD}) / \langle k \rangle_0$  (left panel) and  $(\langle k \rangle_{CBF} - \langle k \rangle_{BHD}) / \langle k \rangle_0$  (right panel) using different immunization algorithms are shown for each of the five empirical networks as a function of the immunization coverage  $f$ . Here,  $\langle k \rangle_0$  is the mean degree of the network before an immunization algorithm is applied. A positive value indicates that BHD is more effective in reducing the mean degree and thus breaking up the network. Results are obtained by averaging over  $10^4$  realizations for each value of  $f$ . (EPS)

**Figure S3 Comparison of duration of epidemic of different immunization algorithms in empirical networks.** The differences in the duration of epidemic between ACQ and BHD (left panel) and between CBF and BHD (right panel) are shown for each of the five empirical networks as a function of the immunization coverage  $f$ . Results are obtained by averaging over  $10^4$  realizations for each value of  $f$ . (EPS)

**Figure S4 Robustness of BHD in networks with noise as modeled by addition and removal of links for 20% immunization coverage.** As in Figure 7, the quantity  $\Delta R_{BHD} = (R_0 - R_{BHD}) / R_0$  is shown as a function of the number of links randomly added to or removed from an empirical network. Results are shown for each of the five empirical networks as labeled. The immunization coverage is  $f = 20\%$ . The SIR

parameters are  $\beta=0.08$  and  $\mu=0.2$ . Results are obtained by averaging over  $10^4$  realizations for each value of added and removed links.

(EPS)

**Figure S5 Effects of starting the search for immunization nodes from different number of steps on the final epidemic ratio.** Results in the main text and figures are obtained by starting the search for immunization nodes two steps after a self-avoiding walk began. Here, results for starting the

search after 1, 2, 3, 4, 5 steps are shown for each of the five empirical networks.

(EPS)

### Text S1 Computational Complexity Analysis.

(PDF)

### Author Contributions

Conceived and designed the experiments: MT PMH YCL YD. Performed the experiments: KG MT HFZ. Analyzed the data: KG MT PMH YCL. Contributed reagents/materials/analysis tools: KG PMH YD. Wrote the paper: KG MT PMH YD YCL.

### References

- Jones JH, Handcock MS (2003) Sexual contacts and epidemic thresholds. *Nature* 423: 605–606.
- Grenfell B, Bjrnstad O (2005) Sexually transmitted diseases: Epidemic cycling and immunity. *Nature* 433: 366–367.
- Keeling M, Rohani P (2007) *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D (2006) Complex networks: Structure and dynamics. *Phys Rep* 424: 175–308.
- Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Phys Rev Lett* 86: 3200–3203.
- Barthelemy M, Barrat A, Pastor-Satorras R, Vespignani A (2004) Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys Rev Lett* 92: 178701.
- Newman MEJ (2005) Threshold effects for two pathogens spreading on a network. *Phys Rev Lett* 95: 108701.
- Bootsma MC, Ferguson NM (2007) The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proc Natl Acad Sci USA* 104: 7588–7593.
- Cauchemez S, Valleron AJ, Boëlle PY, Flahault A, Ferguson NM (2008) Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* 452: 750–754.
- Halloran ME, Ferguson NM, Eubank S, Longini IM, Cummings DAT, et al. (2008) Modeling targeted layered containment of an influenza pandemic in the United States. *Proc Natl Acad Sci USA* 105: 4639–4644.
- Ruan ZY, Tang M, Liu ZH (2012) Epidemic spreading with information-driven vaccination. *Phys Rev E* 86: 036117.
- Zhou J, Xiao G, Cheong SA, Fu X, Wong L, et al. (2012) Epidemic reemergence in adaptive complex networks. *Phys Rev E* 85: 036107.
- Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, et al. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437: 209–214.
- Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, et al. (2006) Strategies for mitigating an influenza pandemic. *Nature* 442: 448–452.
- Wang L, Zhang Y, Huang T, Li X (2012) Estimating the value of containment strategies in delaying the arrival time of an influenza pandemic: A case study of travel restriction and patient isolation. *Phys Rev E* 86: 032901.
- Madar N, Kalisky T, Cohen R, ben Avraham D, Havlin S (2004) Immunization and epidemic dynamics in complex networks. *Eur Phys J B* 38: 269–276.
- Moreno Y, Pastor-Satorras R, Vespignani A (2002) Epidemic outbreaks in complex heterogeneous networks. *Eur Phys J B* 26: 521–529.
- Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, et al. (2010) Identification of influential spreaders in complex networks. *Nat Phys* 6: 888–893.
- Pastor-Satorras R, Vespignani A (2002) Immunization of complex networks. *Phys Rev E* 65: 036104.
- Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, et al. (2005) Infection in social networks: Using network analysis to identify high-risk individuals. *Am J Epidemiol* 162: 1024–1031.
- Yuu Y, Tetsuya Y (2012) A comparative study of community structure based node scores for network immunization. *Lecture Notes in Computer Science* 7669: 328–337.
- Zhang HF, Li KZ, Fu XC, Wang BH (2009) An efficient control strategy of epidemic spreading on scale-free networks. *Chinese Phys Lett* 26: 068901.
- Newman MEJ (2005) A measure of betweenness centrality based on random walks. *Soc Networks* 27: 39–54.
- Cohen R, Havlin S, ben Avraham D (2003) Efficient immunization strategies for computer networks and populations. *Phys Rev Lett* 91: 247901.
- Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99: 7821–7826.
- Lancichinetti A, Kivela M, Saramaki J, Fortunato S (2010) Characterizing the community structure of complex networks. *PLoS ONE* 5: 0011976.
- Traud AL, Kelsic ED, Mucha PJ, Porter SA (2011) Comparing community structure to characteristics in online collegiate social networks. *SIAM review* 53: 526–543.
- Ferrara E (2012) Community structure discovery in facebook. *International Journal of Social Network Mining* 1: 67–90.
- Goncalves B, Perra N, Vespignani A (2011) Modeling users' activity on twitter networks: Validation of dunbar's number. *PLoS ONE* 6: 0022656.
- Liu ZH, Hu B (2005) Epidemic spreading in community networks. *Europhys Lett* 72: 315.
- Huang L, Park K, Lai YC (2006) Information propagation on modular networks. *Phys Rev E* 73: 035103.
- Huang W, Li CG (2007) Epidemic spreading in scale-free networks with community structure. *J Stat Mech* 01: P01014.
- Wu XY, Liu ZH (2008) How community structure influences epidemic spread in social networks. *Physica A* 387: 623–630.
- Chu X, Guan J, Zhang Z, Zhou S (2009) Epidemic spreading in weighted scale-free networks with community structure. *J Stat Mech* 07: P07043.
- Zhou J, Liu ZH (2009) Epidemic spreading in communities with mobile agents. *Physica A* 388: 1228–1236.
- Guimera R, Amaral LAN (2005) Functional cartography of complex metabolic networks. *Nature* 433: 895–900.
- Gong K, Tang M, Yang H, Shangl MS (2011) Variability of contact process in complex networks. *CHAOS* 21: 043130.
- Shu PP, Tang M, Gong K, Liu Y (2012) Effects of weak ties on epidemic predictability in community networks. *CHAOS* 22: 043124.
- Onnela JP, Saramaki J, Hyvonen J, Szabo G, Lazer D, et al. (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci USA* 104: 7332–7336.
- Centola D, Macy M (2007) Complex Contagions and the Weakness of Long Ties. *AJS* 113: 702–734.
- Zhao J, Wu J, Xu K (2010) Weak ties: subtle role of information diffusion in online social networks. *Phys Rev E* 82: 016105.
- Masuda N (2009) Immunization of networks with community structure. *New J Phys* 11: 123018.
- Mihalik A, Csermely P (2011) Heat shock partially dissociates the overlapping modules of the yeast protein-protein interaction network: a systems level model of adaptation. *PLoS Comput Biol* 7: 1002187.
- Yang H, Tang M, Zhang HF (2012) Efficient community-based control strategies in adaptive networks. *New J Phys* 14: 123017.
- Salathe M, Jones JH (2010) Dynamics and control of diseases in networks with community structure. *PLoS Comput Biol* 6: 1000736.
- Chen YP, Paul G, Havlin S, Liljeros F, Stanley HE (2008) Finding a better immunization strategy. *Phys Rev Lett* 101: 058701.
- Ferrara E, Meo P, Fiumara G, Provetti A (2012) The role of strong and weak ties in facebook: a community structure perspective. *arXiv:12030535v1*.
- Guimerà R, Mossa S, Turtschi A, Amaral LAN (2005) The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc Natl Acad Sci USA* 102: 7794.
- Hébert-Dufresne L, Allard A, Young JG, Dubé LJ (2012). Global efficiency of local immunization on complex networks. *arXiv:1208.5768*.
- Fortunato S (2010) Community detection in graphs. *Phys Rep* 486: 75–174.
- Holme P, Kim BJ, Yoon CN, Han SK (2002) Attack vulnerability of complex networks. *Phys Rev E* 65: 056109.
- Lu LY, Zhang YC, Yeung CH, Zhou T (2011) Leaders in social networks, the delicious case. *PLoS ONE* 6: 21202.
- Holme P, Saramaki J (2012) Temporal networks. *Phys Rep* 519: 97–125.
- Keeling MJ, Rohani P (2008) *Modeling infectious diseases in humans and animals*. Princeton University Press.
- Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69: 026113.
- Newman MEJ (2002) Assortative mixing in networks. *Phys Rev Lett* 89: 208701.