PHYSICAL REVIEW E **93**, 032312 (2016)

Role of structural holes in containing spreading processes

Ping Li[*] and Xian Sun
Center for Intelligent and Networked Systems, School of Computer Science, Southwest Petroleum University, Chengdu 610500, China

Kai Zhang
NEC Laboratories America, Inc., 4 Independence Way, Princeton, New Jersey 08540, USA

Jie Zhang
Center for Computational Systems Biology, Fudan University, Shanghai 200433, China

Jürgen Kurths
Potsdam Institute for Climate Impact Research, Potsdam 14415, Germany
and Institute of Physics, Humboldt University of Berlin, Berlin 12489, Germany

(Received 29 October 2015; published 11 March 2016)

Structural holes are channels or paths spanned by a group of indirectly connected nodes and their intermediary in a network. In this work we emphasize the interesting role of structural holes as brokers for information propagation. Based on the distribution of the structural hole numbers associated with each node, we propose a simple yet effective approach for choosing the most influential nodes to immunize in containing the spreading processes. Using a wide spectrum of large real-world networks, we demonstrate that the proposed approach outperforms conventional methods in a remarkable way. In particular, we find that the performance gains of our approach are particularly prominent for networks with high transitivity and assortativity, which verifies the vital role of structural holes in information diffusion on networked systems.

DOI: 10.1103/PhysRevE.93.032312

I. INTRODUCTION

Networks have emerged as an attractive theme in complex system research, due to their universality for depicting a variety of natural and synthetic systems [1,2]. Based on the fact that a complex system consists of many mutually interacting components, the interaction among components and how they intertwine can be captured by a graph, where nodes correspond to individual components and edges to their interactions. Such a network can be thought of as a backbone of the complex system along which signals, information, and contagious entities propagate. In real-world networks, however, undesirable signals can also spread through networks. For example, malicious rumors can spread among individuals, computer worms deluge the internet, and epidemic diseases can infect vulnerable people through contact. Hence, developing effective strategies for preventing the spreading of harmful signals through a network is a research challenge of both theoretical and practical importance.

In general, containing spreading processes, a dual problem to the influence maximization problem (i.e., finding the most influential nodes for information diffusion), can be formulated as an optimization problem that is non-deterministic polynomial-time hard. Therefore, pursuing an exact solution to influence minimization for large-scale networks suffers from combinatorial explosion. Alternatively, a number of efficient approaches are proposed to approximately solve the problem [3–10]. These solutions can be classified into two categories: (i) optimization-based methods; for example, by using the greedy heuristic climbing algorithm, the result of Kempe *et al.* [9] shows that it can approximate the optimum to within a factor of $1 - 1/e$ (where e is the base of the natural logarithm); and (ii) structure-based approaches, typically inspired by the relatedness between the topology and functionality of networks, such as those topology based heuristic methods including [4,5]. Compared with the optimization-based algorithms, heuristics approaches are computationally much more efficient and thus more practical in a diversity of networked systems. Therefore, it is of particular interest to devise efficient methods by leveraging the information of the structures of complex networks.

Since the seminal works of Watts and Strogatz [11] and Barabási and Albert [12], the topological characteristics of networks have been thoroughly explored in sociology [13], biology [14], technology [15], economy [16], etc. One prominent feature is the extremely broad, often scale-free, distribution of degree (defined as the number of immediate neighbors) of their nodes. The property of scale-free degree distribution reflects the allocation of one kind of social capital, from the viewpoint of social network analysis, i.e., a few hub nodes in such networks have a disproportionally large number of interaction partners while the majority of nodes are connected only to just a handful of relations. In other words, to some extent the degree of a node represents the importance of that node in the whole network. Identifying important nodes [17] is of great significance for infection control. In previous works [18–20], heuristic methods based on degrees have been proved to be more efficient than random strategies in selecting candidate nodes to vaccinate. However, node degree only captures network property at low orders, while control information spreading paths of networks obviously requires higher-order information on networks. This is why the degree-based methods are not always optimal.

[*] Corresponding author: dping.li@gmail.com

To improve the performance of heuristic infection control, nontrivial properties of networks based on higher-order connectivity patterns are required. Considering that information often travels across the shortest paths of a network, it is clear that the more paths pass through a node, the more important that node should be. In this sense, betweenness centrality [21] is an appropriate measure for quantifying the importance of a node. While betweenness has wide applications in many networks, it fails to show superiority compared to degree-based methods [3]. The reason lies in the assumptions that information transfers follow shortest paths, while spreading processes expand more randomly in real-world situations. Further, though betweenness centrality considers the indirect influence of high-order neighbors of the nodes, the infection occurring in the ego networks of the infected has no immediate relation to distant nodes. Ego betweenness [22,23] defines node importance at the local level, but as it is highly correlated with global betweenness centrality, the immunization effects resulting from the two betweenness-based strategies are nearly the same.

To find effective indicators by which a set of nodes is chosen to be vaccinated in containing spreading processes, one needs to capture the full characteristics of node influence in general information transfer. Usually it is insufficient to apply only an individual importance indicator to characterize the role of a node in spreading processes. A reasonable solution is to apply multiple indicators to evaluate the importance of a node and identify its role [24]. However, determining the weightings of respective metrics in the composite measure is an open problem. Besides, some metrics such as betweenness centrality are computationally expensive for large-scale networks. Motivated by these considerations, in this work we design a simple metric, which we call structural hole count (SHC), to capture the roles of a node from multiple facets. This metric is then applied to contain spreading processes modeled in a number of real-world networks. Remarkable improvements on the efficiency of the containing strategy are found through numerical simulations. To further understand properties of this indicator, we also inspect its relationship to existing network metrics. Our findings indicate that the structural hole count achieves a combined effect of the degree correlation and clustering coefficients simultaneously in containing spreading processes.

The remainder of this paper is organized as follows. We first introduce a role measurement according to the theory of structural holes [25] in social network analysis in Sec. II and then devise the contain strategy based on the proposed metric in Sec. III. We provide computational experiments to validate the metric on large social networks, showing that the strategy based on our proposed metric significantly outperforms the degree and betweenness centrality based heuristics. We also discuss the performance gains in a range of real networks and find a correlation between the introduced metric and several existing metrics in Sec. IV. Section V gives a summary of this work.

II. MEASURING THE ROLE OF NODES BY STRUCTURAL HOLES

We begin by introducing a metric for quantifying the relative importance of a node in the spreading process of a network, based on the so-called structural holes. The theory of structural holes indicates that some individuals in a network bridge people or clusters of people that are otherwise disconnected. They act as structural hole spanners to fill the holes among the people or group without direct connections. For example, as shown in Fig. 1, there are four structural holes marked by dashed blue arrows associated with the ego node. For each structural hole, the nodes in it will never have the chance to communicate with each other upon the removal of the ego node. Obviously, the ego node plays an important role in the information propagation and the higher the number of structural holes associated with it, the more important the role is. Hereby we leverage the property of structural holes to define an intuitive measure called structural hole count SHC(i) to quantify the node importance, as

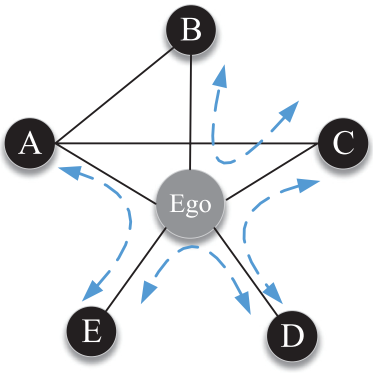

FIG. 1. Schematic diagram of structural holes in an ego network. Dashed lines represent structural holes in the presence of ego.

$$\mathrm{SHC}(i) = \frac{1}{2}\sum_{k,j\in\mathcal{N}_i}(1 - A_{kj}), \tag{1}$$

where $A_{k,j}$ denotes the connectivity between a pair of nodes k and j ($A_{k,j} = 1$ if there is an edge linking node k and node j and $A_{k,j} = 0$ otherwise). Moreover, $\mathcal{N}_i$ represents the set of neighbors of node i. The SHC(i) counts how many structural holes exist in the neighborhood set of node i. In early studies of social network research, Burt used the same index in the studies of ego networks [25], which is called brokage in the context of sociology. However, less attention has been drawn in applying this measurement in network analysis. In this work, we extend the idea to general networks and reveal its important role in spread control.

Here we discuss more properties of the structural hole count. First, it has connections but dissimilarities with ego betweenness in some respects. Both are defined on ego networks, which is centered on a specific node with its personal network as depicted in Fig. 1. Since ego is between two other nodes, if ego lies on the shortest path from one to the other, ego betweenness indicates the percentage of all geodesic paths from neighbor to neighbor passing through ego, while SHC only considers the number of spanned holes by ego.

Second, SHC has an interesting relation with the degree of a node. It can be easily verified that for any node i in a graph, the following relation holds:

$$\mathrm{SHC}(i) = \tfrac{1}{2}\left(d_i^2 - d_i\right) - |G_{\mathcal{N}_i}|, \tag{2}$$

where d_i represents the degree of node i, $\mathcal{N}_i$ denotes the neighbors of i (including the node i itself), $G_{\mathcal{N}_i}$ denotes the subgraph spanned on this set of nodes, and $|\cdot|$ denotes the number of edges in the subgraph. As can be seen, the structural hole count is upper bounded by the squared degrees. This explains the positive correlation between the two quantities, as is demonstrated in Fig. 3, since only nodes with high degrees are likely to have a large structural hole count. However, in the meantime, nodes do exist with high degrees but low structural hole counts: Each of such nodes will be centered by densely interconnected neighbors, leading to a large $|G_{\mathcal{N}_i}|$ that diminishes the structural hole count. Therefore, depending on the actual graph connectivity, the two quantities can demonstrate a significant difference as well.

III. SPREADING CONTAINED IN REAL NETWORKS

In general, the study of spreading suppression is based on information diffusion or epidemic spreading models of networks. The problem of containing the spreading processes is then to reduce the proportion of being infected by blocking as few as possible nodes in a network. In this work, we focus on the susceptible-infected (SI) model [26]. This spreading model usually assumes two possible states for each node, i.e., susceptible and infected. The susceptible state can switch to the infected state with certain probability λ $(0 < \lambda \leqslant 1)$ and then sticks to the infected state for the rest of the time.

Now we formulate the spreading contained problem in a network represented by an undirected graph $G = (V,E)$. Here V and E $(\subset V \times V)$ denote the sets of nodes and links in the network, respectively. The spreading process proceeds from any initially infected node v in the following way. At each discrete time step t, when there exists any susceptible node in the neighborhood of node v, node v tries to contaminate its susceptible neighbors. Then, in the next steps, all currently infected nodes continue to spread the information or disease to their immediate susceptible neighbors until the susceptible become the infected. Eventually, all the nodes are infected if there is no intervention. At this point, the SI model distinguishes itself from the independent-cascade model [9], wherein the latter the infected nodes are only given a single chance to activate each of their susceptible neighbors and therefore the final infection heavily depends on the propagation probability λ.

To measure the severity of contamination in the network G subject to interventions, one can use the average contamination degree (ACD) $c(G,Z)$ defined as

$$c(G,Z) = \frac{1}{|V_Z|} \sum_{v \in V_Z} \sigma(v;G_Z). \tag{3}$$

Here Z is the set of protector nodes that are blocked or immunized, V_Z denotes the residual set of nodes $(V \setminus Z)$, and G_Z denotes the network on V_Z. Here G_Z can be deemed the residual network constructed after blocking the nodes Z in the graph. The term $\sigma(v;G_Z)$ is the number of nodes that are reachable from v. The $|\cdot|$ operator denotes the number of elements in the entity$\cdot$.

The ACD measures, on average, how many victims each node in the residual graph can infect. It can be computed directly given a specific protector set Z. However, this is computationally time consuming for large-scale networks. Instead, we experimentally evaluate the ACD by randomly sampling seed nodes and employing the spreading processes on networks with sufficient repeats.

In order to minimize the spread of contamination in a graph, we propose a heuristic to immunize (or block) those nodes with large structural hole counts. On the one hand, such nodes typically reside in dense regions of the graph (that is, with high degrees), potentially having an impact on many neighbors. On the other hand, neighbors of such nodes only have sparse connections with each other. This means immunization of such nodes will eventually break down almost all the connections within its neighborhood, rendering very few alternative propagation channels left.

In the literature, various solution schemes were also proposed to this problem. For example, it has been shown that removing nodes in order of decreasing degrees (or out-degrees in directed networks) is a successful scheme for preventing the spread of contamination in most real networks [18–20,27]. Some other examples include random walks [28], betweenness criteria [29], and highest-degree neighbors [30] (see [31] for a comprehensive review).

In the following, we compare our method with several popular heuristic methods that are based on the topology of the network. These methods include (i) selecting nodes according their degrees, referred to as the degree-based method; (ii) randomly sampling nodes, referred to as the random scheme; (iii) performing random walk in the network and immunizing every node visited [28], referred to as the random walk; and (iv) selecting the neighboring nodes with largest degree [30], referred to as NDeg. Using several large real networks, we experimentally evaluate the performance in terms of ACD, which is estimated by numerical simulation of spreading dynamics on the network G, as mentioned previously. One of the real networks we employ in our experiments is the collaboration network of Arxiv GrQc (general relativity and quantum cosmology) [32], which comes from the e-print arXiv and depicts scientific collaboration relations between authors whose papers have been submitted to the general relativity and quantum cosmology category. Specifically, if an author i coauthored a paper with author j, there is an undirected edge from i to j. Other real networks such as Astro-Physics (astroph) collaboration networks and High Energy Physics (caHepPh) collaboration network [32] also have been used to verify the validity of the proposed method. It should be noted that the largest connected component will be considered when a network is not strongly connected.

In the experiments, the estimated performance index $c(G,Z)$ is calculated by using the SI model. Although the SI model has the propagation probability λ as a tuning parameter merely determines the time to reach convergence but not the infected area of the spreads. Thus one can take arbitrary positive values less than 1 in the experiment. Here we use $\lambda = 0.05$, but other nonzero values were found to give similar results.

Figure 2 shows the contamination degree of the resulting networks as a function of the number of nodes blocked k, using different schemes as discussed above. We can see that the proposed method outperforms other methods in that the

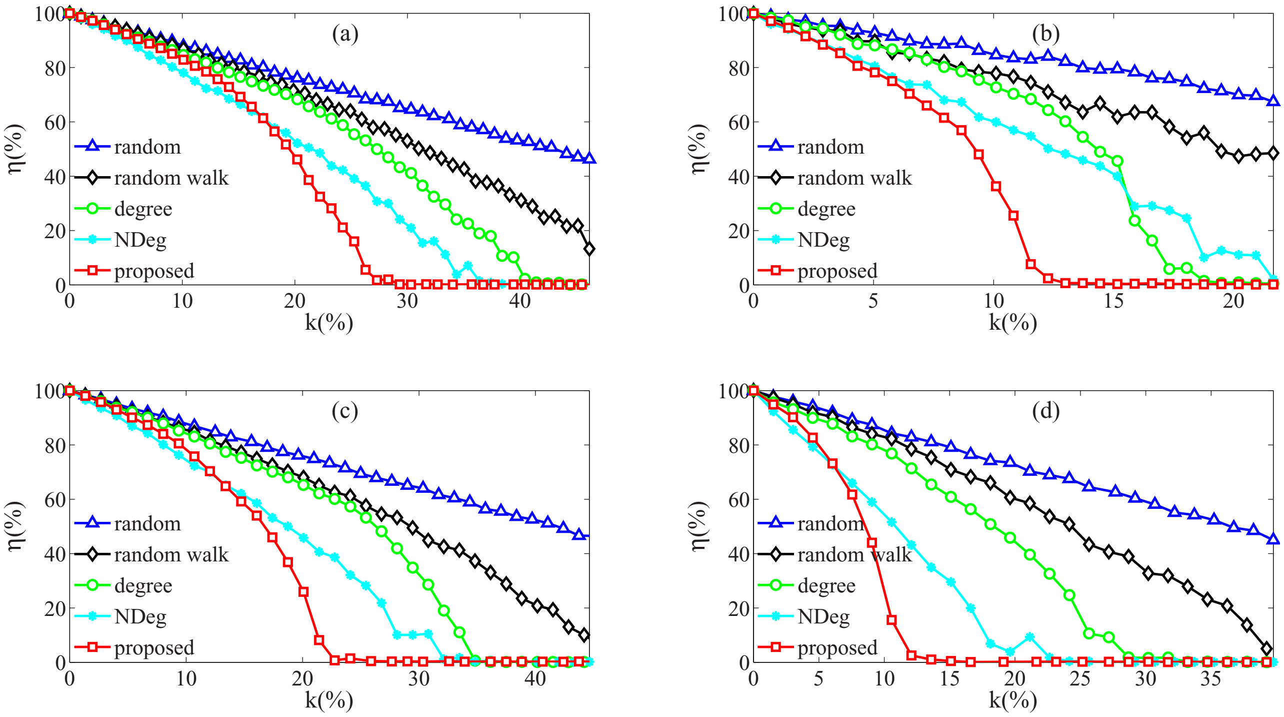

FIG. 2. Estimated average contamination degree as a monotonic function of the number of blocked nodes, compared with degree-based random methods, random walk, and NDeg in (a) astroph, (b) caGrQc, (c) caHepPh, and (d) netPhy networks, respectively.

average contamination degree declines more rapidly over k. More specifically, the expected contamination degree is reduced to almost zero by blocking about 14% of the nodes with the proposed method in the astroph network (Fig. 2). In contrast, by blocking the same number of nodes, the contamination degree is still more than 50% for the degree-based method. Overall, the performances of random schemes (both random sampling and random walk) are inferior; the NDeg method can outperform the degree-based method, indicating the importance of the highest-degree neighbors, and our approach always outperforms other competing methods in these large-scale networks. Another observation is that the performance of the random walk method is inferior to the degree-based method in these networks, which seems to be different from the observations made in [28].

Note that the ultimate contamination degree η may hardly reach zero. By looking into the distribution of vaccinations in the networks, it is found that most of the remaining nodes are surrounded by immunized nodes, indicating that diffusion starting from those nodes can hardly spread out. However, there do exist a few hub nodes that still have some nonimmunized neighbors, which can be infected by the hub nodes, owing to the scale-free property of the degree and structural hole count distributions.

We further compare the sets of target nodes immunized in our method with the degree-based method. The insets in Fig. 3 show the similarities of the two sets of target nodes as a function of the number k of nodes blocked in four real networks. Here the similarity J is quantified by using Jaccard index [33]. It can be observed that for very low immunization rates (k% close to zero), the target nodes from the two methods can be very close. However, when a larger k is needed, the two methods typically target different nodes in the graph, with an overlap ratio that is lower than 50%.

These results imply that some nodes with higher structural hole count but lower degree are crucial for preventing the spreading processes. This can be manifested from the definition of the structural hole count. Structural hole count takes into account not only the connections a node has, but also the connectivity of the neighborhood of this node. In containing the spreading dynamics of a network, blocking the nodes whose neighbors are densely connected with each other can hardly decrease the chance that malicious information spreads from one neighbor to other neighbors, because of the existence of alternative communication channels in the dense ego network. In comparison, the definition of structural hole counts explicitly takes into account the impact of alternative propagation channels. Therefore, the preference is given to those nodes with low connectivities among their neighborhood, rather than those with densely connected neighbors.

IV. DISCUSSION

Through the numerical simulation on a wide spectrum of real-world networks, we have observed that the performance gains of the proposed method over the degree-based method varies from network to network. Specifically, in some networks the proposed method works remarkably better [as Fig. 2(d) shows], while in other cases the improvement can be insignificant. Thus, it is interesting to study what causes the difference in the performance of an intervention method for different networks.

To achieve this, we employ 14 real networks as described in Table I. We adopt some classical measures to characterize these

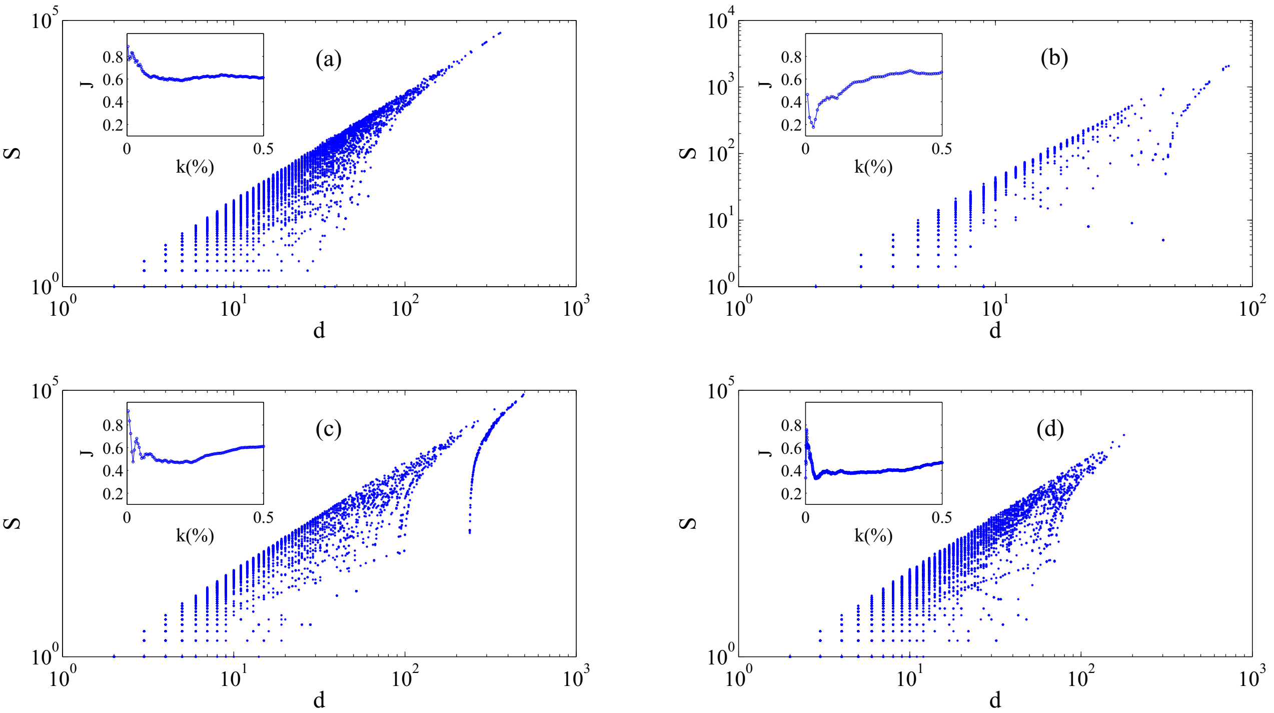

FIG. 3. Structural hole count vs degree for four real networks, i.e., (a) astroph, (b) caGrQc, (c) caHepPh, and (d) netPhy networks. The insets show the similarity of the two sets of blocked nodes corresponding to degree-based and structural-hole-count-based methods.

networks, such as the average shortest path length L, clustering coefficient C [11,43], average degree D, and assortativity coefficient A_c [44]. In addition, we define the performance gains $\varphi(G)$ between the proposed method and the degree-based method, as

$$\varphi(G) = \max_k [c_s(G;k) - c_d(G;k)]. \quad (4)$$

Here k is the ratio of immunized nodes and c_s and c_d are the average contamination degrees corresponding to the proposed and degree-based methods, respectively. Then $\varphi(G)$ is the maximal difference of the performances for the two methods at some k. Clearly, φ indicates the improvement of the proposed method with respect to the degree-based method.

We study the relation between the performance gains and network properties, in particular the clustering coefficient and the assortativity coefficient. Figure 4 shows the performance gains of the proposed method with regard to the two network characteristics. As can be seen, the two network properties can clearly separate those networks with significant and insignificant performance gains. The proposed method seems

TABLE I. Real networks used in the experiments. For disconnected networks, the maximal connected components are taken into consideration.

Network	V	E	$\langle d\rangle$
karate club [34]	34	78	4.59
polblogs [35]	1222	16714	27.36
Celegans [36]	297	2345	15.79
power grid [11]	4941	6594	2.67
hepth citation [37]	5835	13815	4.73
netscience [38]	379	914	4.82
URVemail [39]	1133	5451	9.62
PGP [40]	10680	24316	4.55
caGrQC [37]	4158	13422	6.46
caHepTh [37]	8638	24806	5.74
condmat2005 [41]	36458	171735	9.42
astroph [41]	14845	119652	16.12
as22july06 [42]	22963	48436	4.22
caHepPh [37]	11204	117619	20.99
netPhy [38]	19873	128744	12.96

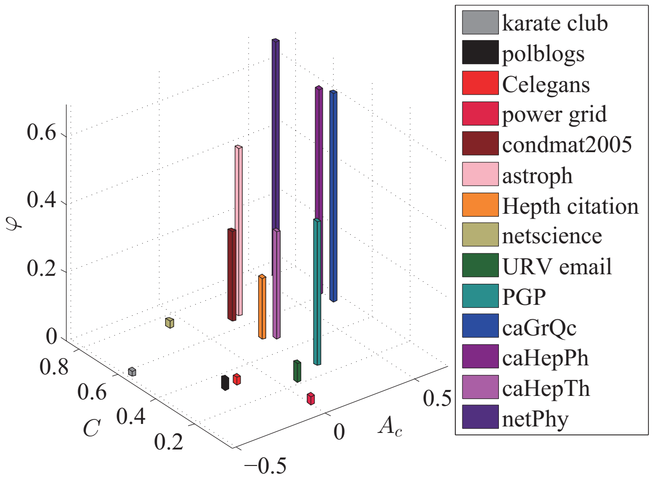

FIG. 4. Performance difference distribution of real networks in two-dimensional topological feature space. Here C denotes the clustering coefficient, A_c represents the assortativity coefficient, and φ is the performance difference.

to have a quite prominent performance gain for networks with high clustering and high assortativity coefficients; in comparison, for networks whose coefficients are both small, the performance gain is insignificant.

The observation indicates that the performance of the proposed method using the structural hole counts is actually adaptive with regard to multiple network characteristics simultaneously. In other words, it equivalently takes into account multiple indicators in containing the spreading processes. We believe this is attributed to the interesting relation between the structural hole counts and various network characteristics. In highly transitive networks, the neighbors of a node are heavily interconnected and form a dense local cluster with their ego, while the nodes that span plenty of structural holes are in the critical paths between different clusters. Thus, intuitively, those bridging nodes should be blocked in suppressing diffusion processes, which coincides with the result demonstrated in [45]. In addition to transitivity, high assortativity of a network implies that highly connected nodes are prone to connecting other highly connected nodes, i.e., hub nodes have hub neighbors. In the targeted containing policy, the degree-based method takes into account the fragility of the hubs in scale-free networks [20]. However, due to the high transitivity and degree-degree correlation, the targeted nodes with high degrees are likely to be densely connected and thus fail to maximize their influence on spreading processes. In contrast, the structural hole count takes into account not only high connections but the type of the connections between the target nodes and the rest nodes. That is, those hubs with weak ties are more important in preventing the infections. Consequently, blocking such nodes will significantly improve the performance of the containing strategy.

Finally, we find that for disassortative networks (i.e., with negative assortativity coefficient), there is no obvious advantage of the proposed method over the degree-based method. In fact, both methods are very efficient in containing the spreading process with a small portion of immunized nodes in disassortative networks. The reason is that disassortative networks often have a scale-free degree distribution. Neighbors of high-degree nodes usually only have a small number of connections and these neighbors are not interconnected directly by virtue of the disassortativity property. In other words, the structural hole counts and the degree are close for those truly important nodes in containing the spreading process. Therefore, high-degree nodes also span the structural holes among the low-degree nodes, which leads to a big overlap between the immunization sets derived from the two methods.

V. CONCLUSION

Developing effective strategies for containing the spread of undesirable propagation through a network is an important topic of both practical and research significance. By leveraging the role of structural holes in shaping the communication channels, we have devised an effective heuristic approach to suppress the spreading processes in a network. Using a variety of real-world networks including several large-scale social networks, we have demonstrated experimentally that the proposed method can significantly outperform the degree-based heuristic for networks with high transitivity and assortativity. We further interpret this success by the correlation of structural holes and weak ties among high-degree nodes. Compared to greedy algorithms, our approach based on the topological characteristics depends merely on local connectivity patterns and is computationally much more attractive for real-world applications.

ACKNOWLEDGMENTS

P.L. was funded by National Natural Science Foundation of China under Grant No. 81373531 and Foundations of SiChuan Educational Committee (Grant No. 13ZB0198). J.Z. was funded by National Science Foundation of China (Grants No. NSFC 61004104 and No. 61104143) and special Funds for Major State Basic Research Projects of China (Grant No. 2015CB856003).

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[2] M. Small, L. Hou, and L. Zhang, Natl. Sci. Rev. **1**, 357 (2014).

[3] E. Khalil, B. Dilkina, and L. Song, Workshop on Frontiers of Network Analysis: Methods, Models, and Applications at NIPS (unpublished).

[4] C. M. Schneider, T. Mihaljev, S. Havlin, and H. J. Herrmann, Phys. Rev. E **84**, 061911 (2011).

[5] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, Physica A **391**, 1777 (2012).

[6] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, Nat. Phys. **6**, 888 (2010).

[7] F. D. Sahneh, F. N. Chowdhury, and C. M. Scoglio, Sci. Rep. **2**, 632 (2012).

[8] W. Chen, Y. Wang, and S. Yang, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, 2009), pp. 199–208.

[9] D. Kempe, J. Kleinberg, and É. Tardos, in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, 2003), pp. 137–146.

[10] A. Lima, V. Pejovic, L. Rossi, M. Musolesi, and M. Gonzalez, arXiv:1504.01316.

[11] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[12] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[13] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge University Press, Cambridge, 2010).

[14] E. Bullmore and O. Sporns, Nat. Rev. Neurosci. **10**, 186 (2009).

[15] P. Erola, S. Gómez, and A. Arenas, in *Proceedings of the 2011 IEEE GLOBECOM Workshops (GC Wkshps)* (IEEE, Piscataway, 2011), pp. 95–99.

[16] M. O. Jackson *et al.*, *Social and Economic Networks* (Princeton University Press, Princeton, 2008), Vol. 3.

[17] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. Lett. **97**, 094102 (2006).

[18] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).

[19] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Comput. Networks **33**, 309 (2000).

[20] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).

[21] M. Barthelemy, Eur. Phys. J. B **38**, 163 (2004).

[22] L. C. Freeman, Math. Soc. Sci. **3**, 291 (1982).

[23] M. Everett and S. P. Borgatti, Soc. Networks **27**, 31 (2005).

[24] P. Hu, W. Fan, and S. Mei, Physica A **429**, 169 (2015).

[25] R. S. Burt, *Structural Holes: The Social Structure of Competition* (Harvard University Press, Cambridge, 2009).

[26] L. J. Allen, Math. Biosci. **124**, 83 (1994).

[27] M. E. J. Newman, S. Forrest, and J. Balthrop, Phys. Rev. E **66**, 035101 (2002).

[28] K. Hu and Y. Tang, Chin. Phys. **15**, 2782 (2006).

[29] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, Phys. Rev. E **65**, 056109 (2002).

[30] P. Holme, Europhys. Lett. **68**, 908 (2004).

[31] R. Pastor-Satorras, C. Ccastellano, P. Van Mieghem, and A. Vespignani, Rev. Mod. Phys. **87**, 925 (2015).

[32] J. Leskovec, J. Kleinberg, and C. Faloutsos, ACM Trans. Knowledge Discovery Data **1**, 2 (2007).

[33] T. Pang-Ning, M. Steinbach, and V. Kumar, *Library of Congress* (Addison-Wesley, Boston, 2006), p. 74.

[34] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).

[35] L. A. Adamic and N. Glance, in *Proceedings of the Third International Workshop on Link Discovery* (ACM, New York, 2005), pp. 36–43.

[36] J. Duch and A. Arenas, Phys. Rev. E **72**, 027104 (2005).

[37] J. Leskovec and A. Krevl, SNAP Datasets: Stanford large network dataset collection, http://snap.stanford.edu/data (2014).

[38] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

[39] R. Guimera, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103 (2003).

[40] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, Phys. Rev. E **70**, 056122 (2004).

[41] M. E. J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001).

[42] M. E. J. Newman, http://www-personal.umich.edu/˜mejn/netdata/ (2006).

[43] P. W. Holland and S. Leinhardt, Comp. Group Stud. **2**, 107 (1971).

[44] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[45] K. S. Andreas I. Reppas, and C. I. Siettos, Virulence **3**, 146 (2012).