

RESEARCH ARTICLE

A Novel Top- k Strategy for Influence Maximization in Complex Networks with Community Structure

Jia-Lin He^{1,2}, Yan Fu^{1,2}, Duan-Bing Chen^{1,2*}

1 Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China, **2** Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

* dbchen@uestc.edu.cn



OPEN ACCESS

Citation: He J-L, Fu Y, Chen D-B (2015) A Novel Top- k Strategy for Influence Maximization in Complex Networks with Community Structure. PLoS ONE 10 (12): e0145283. doi:10.1371/journal.pone.0145283

Editor: Lidia Adriana Braunstein, IFIMAR, UNMdP-CONICET, ARGENTINA

Received: June 16, 2015

Accepted: December 2, 2015

Published: December 18, 2015

Copyright: © 2015 He et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data set used in this paper was collected from the website: <http://konect.uni-koblenz.de/networks/>.

Funding: This work is partially supported by the National Natural Science Foundation of China (<http://isisn.nsf.gov.cn/egrantindex/funcindex/prjsearch-list#>) under Grant No. 61433014, by the National High Technology Research and Development Program under Grant No. 2015AA7115089 and by the Fundamental Research Funds for the Central Universities (<http://isisn.nsf.gov.cn/egrantindex/funcindex/prjsearch-list#>) under Grant No. ZYGX2014Z002. The funders had no role in study

Abstract

In complex networks, it is of great theoretical and practical significance to identify a set of critical spreaders which help to control the spreading process. Some classic methods are proposed to identify multiple spreaders. However, they sometimes have limitations for the networks with community structure because many chosen spreaders may be clustered in a community. In this paper, we suggest a novel method to identify multiple spreaders from communities in a balanced way. The network is first divided into a great many super nodes and then k spreaders are selected from these super nodes. Experimental results on real and synthetic networks with community structure show that our method outperforms the classic methods for degree centrality, k -core and ClusterRank in most cases.

Introduction

Spreading process is one of the fundamental processes taking place in complex networks [1–5]. It has been applied in many fields, such as information diffusion [6], disease propagation [4], cascade failure [7], etc. Identifying a set of critical spreaders is an important issue in spreading process [8–11]. For example, in August 2003, three burned power lines in Northern Ohio brought about serious disaster that the entire US Northeast and parts of Canada were plunged into darkness. If the vulnerable regions in power-grid network are known well in advance, we could take some measures to protect them. So a set of critical spreaders is crucial for developing efficient strategies to control the spreading process in complex networks.

In the past years, some special methods have been proposed to identify multiple spreaders. Kempe et al. [12] presented a hill-climbing strategy to choose k spreaders. They demonstrated that the greedy strategy achieves an approximation guarantee of $(1-1/e)$ where e is the base of the natural logarithm. Narayanam et al. [13] proposed a SPIN heuristic algorithm for the top- k nodes problem. To compute the Shapley values required by the SPIN algorithm, they use a simple sampling technique to obtain a computationally efficient scheme. Zhao et al. [14] made an attempt to find effective multiple spreaders in complex networks by generalizing the idea of the

design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

coloring problem in graph theory to complex networks. In their method, the nodes with the same color are sorted into an independent set. Then, for a given centrality, the nodes with the highest centrality in an independent set are chosen as multiple spreaders. Chen et al. [15] proposed degree discount heuristics, which nearly match the performance of the greedy methods for the IC model, while also improve upon the pure degree heuristic in other cascade models. Zhang et al. [16] proposed a novel method for identifying influential nodes in complex networks with community structure. The method uses the information transfer probability between any pair of nodes and the k -medoid clustering algorithm.

There are two benchmark methods for the identification of multiple spreaders in complex networks. The first one chooses the top k influential nodes as spreaders according to a centrality index [17–29]. Although the method is very simple, most of these k spreaders may be clustered in a community. The second one chooses k unconnected spreaders according to a centrality index. However, many spreaders may still locate in a community. In this paper, we suggest a novel method which disperses k spreaders. A network is first divide into a great many super nodes and then k spreaders are chosen from these super nodes according to a centrality index. If a super node includes one spreader, the nodes, which have edges incident to the super node, can not be selected as spreaders any more. The SIR model is used to test the performance of our method. Experimental results on real and synthetic networks with community structure show that our method outperforms the benchmark methods for degree centrality, k -core and ClusterRank in most cases.

Materials and Methods

Super Node

Loosely speaking, a community is a subgraph of a network whose nodes are more tightly connected with each other than with nodes outside the subgraph. Usually, a community exhibits hierarchical organization, that is, it can contain groups of sub-communities, and so forth over multiple scales. [30]. The community hierarchy can be found by Blondel method [31], which is composed of two steps. In the first step, each community adjusts their nodes according to the increment of modularity. In the second step, each community is replaced by a new node called “super node”. The two steps are repeated until the modularity can not be improved. In this paper, to obtain a great many communities, the two steps are iterated only once.

Red-Black Tree

The red-black tree [32] is a type of binary search tree where costs are guaranteed to be logarithmic, no matter what sequence of keys is used to construct them. In the tree, each node is either red or black. It has perfect black balance, i.e., every path from the root to a null link contains the same number of black nodes. The average length of a path from the root to a node in a red-black tree with n nodes is approximately equal to $\log n$. So in a red-black tree, searching operation, insertion operation or ranking operation takes only logarithmic time in the worst case.

Spreader Identification

All super nodes are stored in a red-black tree. For a super node, the key is its id and the values contain its size and its nodes. Besides, it contains a state variable which indicates whether the super node is visited. We first take a non-visited super node with maximal size from the red-black tree. Then we select the most influential node from the super node as a spreader according to a centrality index. Similarly, we take the next super node from the red-black tree and select the most influential node as a spreader, which has no edges incident to the super nodes

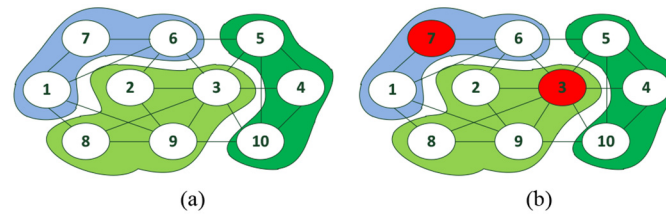


Fig 1. The spreader identification process of our method. (a) A toy network with 10 nodes and 3 super-nodes; (b) two spreaders identified by our method.

doi:10.1371/journal.pone.0145283.g001

which have already contained spreaders. If all super nodes are visited and the number of chosen spreaders is not enough, we restart to visit all super nodes in the descending order of their size and choose the remaining spreaders. The process is repeated until k spreaders are found. In practice, the number of super nodes is far more than that of the spreaders. So k spreaders can be always identified in the first sweep.

In Fig 1, we use a toy network with 10 nodes and 3 super nodes to illustrate our method. Two spreaders will be chosen from the network and degree centrality is used to measure the influence of each node. As shown in Fig 1(a), three super nodes are represented by three different colors respectively. For the biggest super node (2,3,8,9), node 3 is the most influential node and is chosen as a spreader. Nodes 1,4,5,6 and 10 can not be selected as spreaders because they have at least one edge incident to the super node (2,3,8,9). Node 7 is chosen as the second spreader and the final result is shown in Fig 1(b).

Computational Complexity

The computational complexity of our method is analyzed as follows. The super nodes can be found in $O(m)$ time by using the Blondel method, where m is the number of edges in network. Since an insertion operation in a red-black tree with r super nodes takes $O(\log r)$ time, so the construction of a red-black tree with l super nodes takes $O(\log(l-1)!) < O(l \log l)$ time. A searching operation in a red-black tree is guaranteed to visit at most $\log l$ nodes, so k visits totally take $O(k \log l)$ time. Finally, identifying a spreader in a super node takes $O(s)$ in the worse case, where s is the size of the super node and identifying k spreaders totally take $O(n)$ time. So the total running time of our method is $O(m + n + (k + l) \log l)$.

Results

We simulate the spreading process in a network by using the SIR model [33] which has been extensively studied. In the SIR model, each node has one of three states (Susceptible, Infected and Recovered) at each time step. An infected node randomly contacts a neighbor node and transmits the disease to it with a probability μ if the neighbor node is a susceptible one. At the same time, an infected node will be recovered with a probability β . The effective spreading rate λ is defined as μ/β . When there is no infected nodes in a network, the spreading process stops.

Real Networks

The performance of our method is evaluated on three real networks, including Gowalla, Dblp and Youtube networks. Gowalla network [34] contains user-user friendship relations. Nodes represent users and an edge indicates a friendship between two users. Dblp network [35] is a co-authorship network from computer science bibliography. Nodes represent authors and an edge between two nodes exists if two corresponding authors have published at least one paper

Table 1. The topological properties of three real networks, including the number of nodes, the number of edges, the number of super nodes, average degree ($\langle k \rangle$), modularity (Q), mean squared degree ($\langle k^2 \rangle$) clustering coefficient (cc), power law exponent (α) and maximal k -core value (k -core).

network	#nodes	#edges	#super-nodes	$\langle k \rangle$	Q	$\langle k^2 \rangle$	cc	α	k -core
Gowalla	196591	950327	21954	9.67	0.63	2964.03	0.024	2.65	51
Dblp	317080	1049866	63479	6.62	0.58	144.01	0.306	3.26	113
Youtube	1134890	2987624	172928	5.26	0.61	2603.72	0.006	2.14	51

doi:10.1371/journal.pone.0145283.t001

together. Youtube network [35] is a social network from a video-sharing web site. Users form friendships with each other and users can create groups in which other users can join. In the network, nodes represent users and an edge between two nodes indicates a friendship. The detailed information of the three real networks is listed in Table 1.

We compare our method (labeled as super-node) with two benchmark methods on three real networks. The first method (labeled as influential-node) chooses the top k influential nodes as spreaders according to a centrality index. The second method (labeled as disperse-node) first computes a ranking list of nodes based on a centrality index and then selects k unconnected spreaders from the ranking list. Three centrality indices, i.e., degree centrality, k -core and ClusterRank, are chosen to measure the influence of each node in network.

From Fig 2, it can be seen that both our method and the disperse-node method outperform the influential-node method greatly in most cases. So the following analysis only involves our method and the disperse-node method. To quantify the performance of two methods, we define an index called “growth ratio”,

$$growth\ ratio = \frac{P_{our\ method} - P_{other\ method}}{P_{other\ method}}, \tag{1}$$

where $p_{our\ method}$ is the proportion of infected nodes in a network for our method and $p_{other\ method}$ for benchmark method. Fig 3 shows that our method influences a greater scope than the disperse-node method in most cases. It is noted that the growth ratio is related to network structure. All growth ratios for Dblp network are low and most of them are less than 10%.

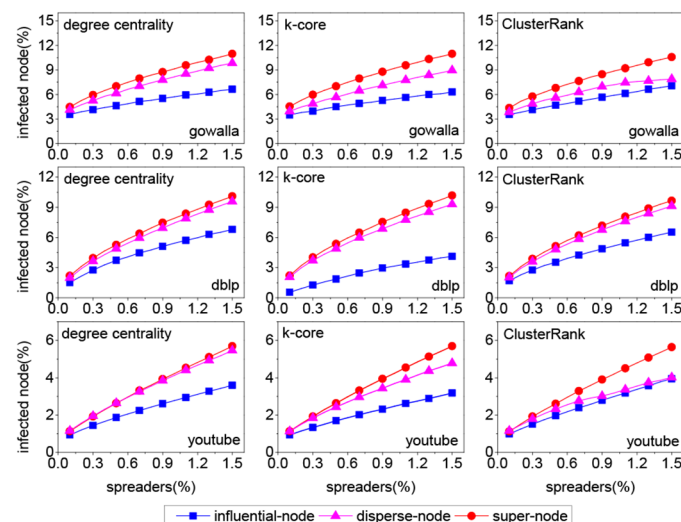


Fig 2. The influence scope with different proportions of spreader on three real networks, where $\lambda = 1.5$, $\beta = 1/\langle k \rangle$. Each data point is obtained by averaging over 200 independent runs.

doi:10.1371/journal.pone.0145283.g002

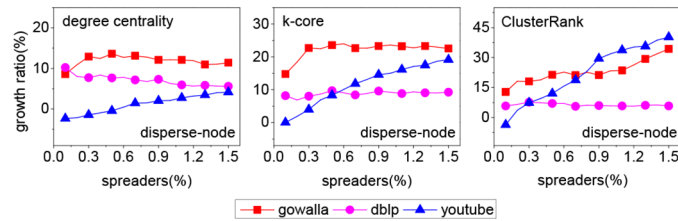


Fig 3. The growth ratio on three real networks for three centrality indices.

doi:10.1371/journal.pone.0145283.g003

However, for the other two networks, most of growth ratios are above 10% and the maximum is more than 30%. Meanwhile, the growth ratio has also to do with centrality index. For degree centrality, the growth ratios are less than 20% on three networks. For *k*-core, most of growth ratios are more than 20% on Gowalla network. For ClusterRank, most of growth ratios are more than 30% on Gowalla and Youtube networks.

To further evaluate the performance of our method, we compare it with the *k*-medoid method [16], which also chooses *k* spreaders from communities. In the *k*-medoid method, each edge(*u*, *v*) is randomly designated either “open” with probability β_{uv} or “closed” with probability $1 - \beta_{uv}$ independently. The β_{uv} is defined as

$$\beta_{uv} = 1 - (1 - \beta)^{w_{uv}}, \tag{2}$$

where w_{uv} is the weight of edge(*u*, *v*) and β is a designated propagation probability. For two nodes *p* and *q*, if there is at least a path between them which is composed of “open” edges, $\omega(p, q) = 1$, otherwise 0. Then the element m_{pq} of information transfer probability matrix *M* is defined as

$$m_{pq} = \frac{1}{N} \sum_{i=1}^N \omega(p, q), \tag{3}$$

where *N* is the number of sampling. The network is first divided into *k* communities based on *M* and then *k* medoids are chosen as *k* spreaders. In the *k*-medoid method, the time complexity of each iteration is $O(k(n - k)^2)$, where *n* is the number of nodes in network. So the method is very time consuming.

Because of high time complexity, the *k*-medoid method can not be applied to Gowalla, Dblp and Youtube networks. So two small real networks, i.e., karate [36] and football networks [37], are used in this experiment. Karate network reflects the social relations of a karate club in an American university. Its nodes represent club members, and an edge indicates social communication between two club members. It includes 34 nodes and 78 edges. Football network is the match network of American football games between Division IA colleges during regular season Fall 2000. Its nodes represent teams, and an edge indicates that a match is played between the two corresponding teams. It contains 115 nodes and 613 edges. The detailed information of the two real networks is described in Table 2.

Table 2. The topological properties of two real networks, including the number of nodes, the number of edges, the number of super nodes, average degree (<*k*>) and modularity (Q).

network	#nodes	#edges	#super-nodes	< <i>k</i> >	Q
Karate	34	78	7	4.59	0.37
Football	115	613	13	10.67	0.58

doi:10.1371/journal.pone.0145283.t002

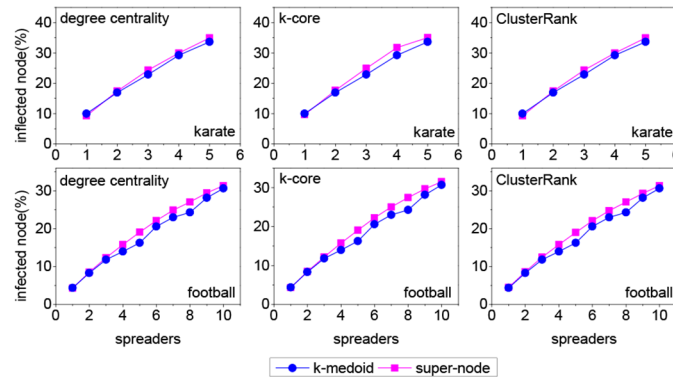


Fig 4. The comparisons between our method and the *k*-medoid method on karate and football networks, where $\lambda = 1.1$, $\beta = 1/\langle k \rangle$. Each data point is obtained by averaging over 100000 independent runs.

doi:10.1371/journal.pone.0145283.g004

From Fig 4, it can be seen that two methods have approximate performance. However, in most cases, our method outperforms the *k*-medoid method slightly. Besides, compared with the *k*-medoid method, our method has two advantages. First, the *k*-medoid method must divide a network into *k* communities to choose *k* spreaders. However, the *k* communities may not meet the community definition, that is, the nodes are denser within communities than across. For our method, the detected communities correspond to the real communities in network because the Blondel method is employed. Second, it is difficult to apply the *k*-medoid method to large networks because of high time complexity. Conversely, our method can choose *k* spreaders quickly in large networks because of low time complexity.

Synthetic Network

We also test the performance of our method on three synthetic scale-free networks which are generated by LFR model [38]. In the LFR model, both the degree and the community size distributions are power laws, with exponents α and β , respectively. In our experiment, three synthetic networks have the same parameters α and β , which are set to 2.5 and 2.5 respectively. The only difference for three synthetic networks is the mixing parameter μ , which is set to 0.1, 0.3 and 0.5 respectively. The detailed information of the three synthetic networks is described in Table 3.

From Fig 5, it can be seen that our method outperforms two benchmark methods in most cases. The corresponding growth ratio is shown in Fig 6. In most cases, the growth ratio is the highest for the LFR1 network and the lowest for the LFR3 network. Interestingly, the modularity of the LFR1 network is the highest and that of the LFR3 network is the lowest, as shown in Table 3. So the growth ratio is proportional to the modularity of network in most cases. The reason can be explained from two aspects, i.e., the structure of super node and the dispersion degree of *k* spreaders. First, the higher the modularity is, the denser the structure of the super

Table 3. The topological properties of three synthetic networks, including the number of nodes, the number of edges, minimum degree (k_{min}), average degree ($\langle k \rangle$), the number of super nodes and modularity (Q).

network	#nodes	#edges	k_{min}	$\langle k \rangle$	α	β	μ	#super-nodes	Q
LFR1	10000	65338	5	13.07	2.5	2.5	0.1	154	0.74
LFR2	10000	65338	5	13.07	2.5	2.5	0.3	225	0.57
LFR3	10000	65338	5	13.07	2.5	2.5	0.5	716	0.35

doi:10.1371/journal.pone.0145283.t003

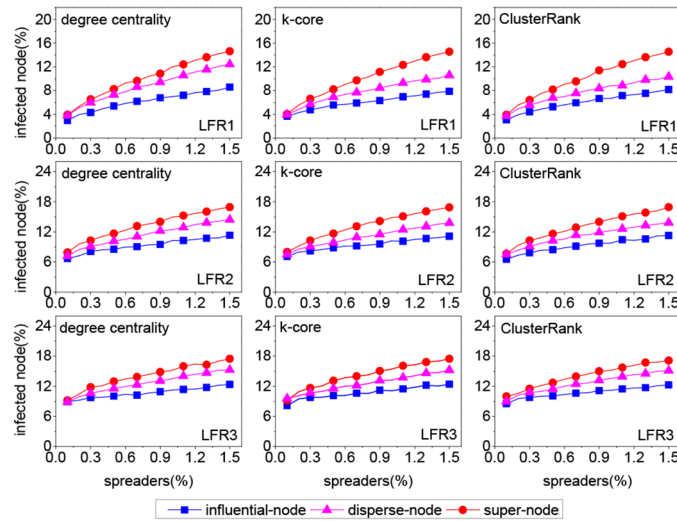


Fig 5. The influence scope with different proportions of spreader on three synthetic networks, where $\lambda = 1.5$, $\beta = 1/k$. Each data point is obtained by averaging over 200 independent runs.

doi:10.1371/journal.pone.0145283.g005

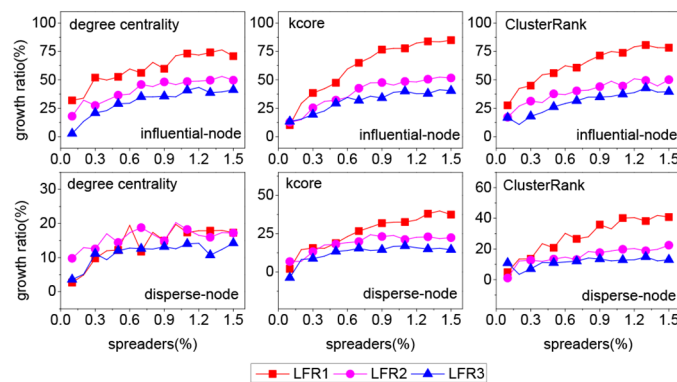


Fig 6. The grow ratios on three synthetic networks.

doi:10.1371/journal.pone.0145283.g006

node is. If two or more spreaders locate in a dense super node, they have many common neighbors. Once the common neighbors are infected, these spreaders have less chances to contact susceptible nodes at each time step. Second, to quantify the dispersion degree of k spreaders, we define an index named “coverage ratio”,

$$coverage\ ratio = \frac{\#super - node'}{\#super - node}, \tag{4}$$

where #super-node is the number of all super nodes in network and #super-node' is the number of super nodes which contain at least one spreader in network. As shown in Fig 7, the coverage ratio of our method is higher than that of two benchmark methods. Take the LFR1 network for example, the coverage ratio is more than 80% for our method, less than 50% for the disperse-node method and less than 20% for the influential-node method. So compared with two benchmark methods, our k spreaders are more disperse. In fact, for our method, a super node usually contains at most one spreader because the number of super nodes is far more than that of the spreaders. However, for two benchmark methods, many super nodes

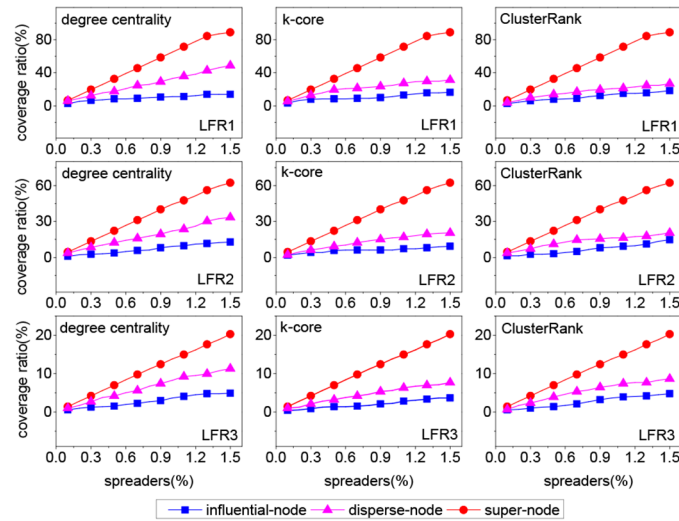


Fig 7. The coverage ratios on three synthetic networks.

doi:10.1371/journal.pone.0145283.g007

contain two or more spreaders. From the above analysis, our method is suitable for the networks with obvious community structure.

Discussion

In this paper, we suggest a novel top- k strategy which chooses multiple spreaders from communities. In our method, the network is first divided into many super nodes and then k spreaders are selected from these super nodes. If a super node contains one spreader, the nodes, which have at least one edge incident to the super node, are not chosen as spreaders any more. In practice, the number of super nodes is far more than that of spreaders, so a super node usually contains at most one spreader.

The performance of our method is evaluated on real and synthetic networks with community structure. On three large real networks, our method outperforms two benchmark methods in most cases. The growth ratio is not only related to network structure but also has to do with centrality index. On two small real networks, our method outperforms the k -medoid method slightly in most cases. Compared with the k -medoid, our method has two advantages. First, the detected communities correspond to the real communities in network. Second, the time complexity is low. On three synthetic scale-free networks, our method still outperforms two benchmark methods in most cases. Compared with two benchmark methods, our method has more chances to contact susceptible nodes on the synthetic network with high modularity.

There are two open issues needing further study in the future. First, the performance of our method is related to centrality index. So how the centrality index affects the identification of multiple spreaders should be studied. Second, with the available of temporal data in recent years, the spreading process in temporal networks has caused great concern [39, 40]. So the further research on the spreader identification in temporal networks is needed.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China under Grant No. 61433014, by the National High Technology Research and Development Program under Grant No. 2015AA7115089 and by the Fundamental Research Funds for the Central Universities under Grant No. ZYGX2014Z002.

Author Contributions

Conceived and designed the experiments: JLH DBC YF. Performed the experiments: JLH. Analyzed the data: JLH DBC. Contributed reagents/materials/analysis tools: JLH DBC YF. Wrote the paper: JLH YF.

References

1. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data[J]. *Nature*, 2009, 457(7232): 1012–1014. doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634) PMID: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)
2. Wang P, González MC, Hidalgo CA, Barabási A L. Understanding the spreading patterns of mobile phone viruses[J]. *Science*, 2009, 324(5930): 1071–1076. doi: [10.1126/science.1167053](https://doi.org/10.1126/science.1167053) PMID: [19342553](https://pubmed.ncbi.nlm.nih.gov/19342553/)
3. Centola D. The spread of behavior in an online social network experiment[J]. *Science*, 2010, 329(5996): 1194–1197. doi: [10.1126/science.1185231](https://doi.org/10.1126/science.1185231) PMID: [20813952](https://pubmed.ncbi.nlm.nih.gov/20813952/)
4. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals[M]. Princeton University Press, 2008.
5. Aral S, Walker D. Identifying influential and susceptible members of social networks[J]. *Science*, 2012, 337(6092): 337–341. doi: [10.1126/science.1215842](https://doi.org/10.1126/science.1215842) PMID: [22722253](https://pubmed.ncbi.nlm.nih.gov/22722253/)
6. Goldenberg J, Libai B, Muller E. Talk of the network: A complex systems look at the underlying process of word-of-mouth[J]. *Marketing Letters*, 2001, 12(3): 211–223. doi: [10.1023/A:1011122126881](https://doi.org/10.1023/A:1011122126881)
7. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. Catastrophic cascade of failures in interdependent networks[J]. *Nature*, 2010, 464(7291): 1025–1028. doi: [10.1038/nature08932](https://doi.org/10.1038/nature08932) PMID: [20393559](https://pubmed.ncbi.nlm.nih.gov/20393559/)
8. Albert R, Jeong H, Barabási A L. Error and attack tolerance of complex networks[J]. *Nature*, 2000, 406(6794): 378–382. doi: [10.1038/35019019](https://doi.org/10.1038/35019019) PMID: [10935628](https://pubmed.ncbi.nlm.nih.gov/10935628/)
9. Callaway DS, Newman ME, Strogatz SH, Watts DJ. Network robustness and fragility: Percolation on random graphs[J]. *Physical Review Letters*, 2000, 85(25): 5468. doi: [10.1103/PhysRevLett.85.5468](https://doi.org/10.1103/PhysRevLett.85.5468) PMID: [11136023](https://pubmed.ncbi.nlm.nih.gov/11136023/)
10. Weng J, Lim EP, Jiang J, He Q. Twitterrank: finding topic-sensitive influential twitterers[C]//Proceedings of the Third ACM International Conference on Web Search and Data Mining. ACM, 2010: 261–270.
11. Vitali S, Glattfelder JB, Battiston S. The network of global corporate control[J]. *PLoS ONE*, 2011, 6(10): e25995. doi: [10.1371/journal.pone.0025995](https://doi.org/10.1371/journal.pone.0025995) PMID: [22046252](https://pubmed.ncbi.nlm.nih.gov/22046252/)
12. Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network[C]//Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2003: 137–146.
13. Narayanam R, Narahari Y. A shapley value-based approach to discover influential nodes in social networks[J]. *IEEE Transactions on Automation Science and Engineering*, 2011, 1(8): 130–147. doi: [10.1109/TASE.2010.2052042](https://doi.org/10.1109/TASE.2010.2052042)
14. Zhao XY, Huang B, Tang M, Zhang HF, Chen DB. Identifying effective multiple spreaders by coloring complex networks[J]. *EPL (Europhysics Letters)*, 2014, 108(6): 68005. doi: [10.1209/0295-5075/108/68005](https://doi.org/10.1209/0295-5075/108/68005)
15. Chen W, Wang Y, Yang S. Efficient influence maximization in social networks[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2009: 199–208.
16. Zhang X, Zhu J, Wang Q, Zhao H. Identifying influential nodes in complex networks with community structure[J]. *Knowledge-Based Systems*, 2013, 42: 74–84. doi: [10.1016/j.knosys.2013.01.017](https://doi.org/10.1016/j.knosys.2013.01.017)
17. Bonacich P. Factoring and weighting approaches to status scores and clique identification[J]. *Journal of Mathematical Sociology*, 1972, 2(1): 113–120. doi: [10.1080/0022250X.1972.9989806](https://doi.org/10.1080/0022250X.1972.9989806)
18. Bavelas A. Communication patterns in task-oriented groups[J]. *The Journal of the Acoustical Society of America*, 1950, 22(6): 725–730. doi: [10.1121/1.1906679](https://doi.org/10.1121/1.1906679)
19. Sabidussi G. The centrality index of a graph[J]. *Psychometrika*, 1966, 31(4): 581–603. doi: [10.1007/BF02289527](https://doi.org/10.1007/BF02289527) PMID: [5232444](https://pubmed.ncbi.nlm.nih.gov/5232444/)
20. Freeman LC. A set of measures of centrality based on betweenness[J]. *Sociometry*, 1977, 40(1): 35–41. doi: [10.2307/3033543](https://doi.org/10.2307/3033543)
21. Kitsak M, Gallos L, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of influential spreaders in complex networks[J]. *Nature Physics*, 2010, 6(11): 888–893. doi: [10.1038/nphys1746](https://doi.org/10.1038/nphys1746)

22. Bae J, Kim S. Identifying and ranking influential spreaders in complex networks by neighborhood core-ness[J]. *Physica A: Statistical Mechanics and its Applications*, 2014, 395: 549–559. doi: [10.1016/j.physa.2013.10.047](https://doi.org/10.1016/j.physa.2013.10.047)
23. Brin S, Page L. Reprint of: The anatomy of a large-scale hypertextual web search engine[J]. *Computer Networks*, 2012, 56(18): 3825–3833. doi: [10.1016/j.comnet.2012.10.007](https://doi.org/10.1016/j.comnet.2012.10.007)
24. Lü L, Zhang YC, Yeung CH, Zhou T. Leaders in social networks, the delicious case[J]. *PLoS ONE*, 2011, 6(6): e21202. doi: [10.1371/journal.pone.0021202](https://doi.org/10.1371/journal.pone.0021202) PMID: [21738620](https://pubmed.ncbi.nlm.nih.gov/21738620/)
25. Chen D, Lü L, Shang MS, Zhang YC, Zhou T. Identifying influential nodes in complex networks[J]. *Physica A: Statistical Mechanics and its Applications*, 2012, 391(4): 1777–1787. doi: [10.1016/j.physa.2011.09.017](https://doi.org/10.1016/j.physa.2011.09.017)
26. Chen DB, Gao H, Lü L, Zhou T. Identifying influential nodes in large-scale directed networks: The role of clustering[J]. *PLoS ONE*, 2013, 8(10): e77455. doi: [10.1371/journal.pone.0077455](https://doi.org/10.1371/journal.pone.0077455) PMID: [24204833](https://pubmed.ncbi.nlm.nih.gov/24204833/)
27. Ren ZM, Zeng A, Chen DB, Liao H, Liu JG. Iterative resource allocation for ranking spreaders in complex networks[J]. *EPL (Europhysics Letters)*, 2014, 106(4): 48005. doi: [10.1209/0295-5075/106/48005](https://doi.org/10.1209/0295-5075/106/48005)
28. Chen DB, Xiao R, Zeng A, Zhang YC. Path diversity improves the identification of influential spreaders [J]. *EPL (Europhysics Letters)*, 2013, 104(6): 68006. doi: [10.1209/0295-5075/104/68006](https://doi.org/10.1209/0295-5075/104/68006)
29. Pu J, Chen X, Wei D, Liu Q, Deng Y. Identifying influential nodes based on local dimension[J]. *EPL (Europhysics Letters)*, 2014, 107(1): 10010. doi: [10.1209/0295-5075/107/10010](https://doi.org/10.1209/0295-5075/107/10010)
30. Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks[J]. *New Journal of Physics*, 2009, 11(3): 033015. doi: [10.1088/1367-2630/11/3/033015](https://doi.org/10.1088/1367-2630/11/3/033015)
31. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks [J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, 2008(10): P10008. doi: [10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008)
32. Cormen TH. *Introduction to algorithms*[M]. MIT press, 2009.
33. Yang R, Wang BH, Ren J, Bai WJ, Shi ZW, Wang WX, et al. Epidemic spreading on heterogeneous networks with identical infectivity[J]. *Physics Letters A*, 2007, 364(3): 189–193. doi: [10.1016/j.physleta.2006.12.021](https://doi.org/10.1016/j.physleta.2006.12.021)
34. Cho E, Myers SA, Leskovec J. Friendship and mobility: user movement in location-based social networks[C]//*Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2011: 1082–1090.
35. Yang J, Leskovec J. Defining and evaluating network communities based on ground-truth[J]. *Knowledge and Information Systems*, 2015, 42(1): 181–213. doi: [10.1007/s10115-013-0693-z](https://doi.org/10.1007/s10115-013-0693-z)
36. Zachary WW. An information flow model for conflict and fission in small groups[J]. *Journal of Anthropological Research*, 1977, 33(4): 452–473.
37. Girvan M, Newman M E J. Community structure in social and biological networks[J]. *Proceedings of the National Academy of Sciences*, 2002, 99(12): 7821–7826. doi: [10.1073/pnas.122653799](https://doi.org/10.1073/pnas.122653799)
38. Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms [J]. *Physical Review E*, 2008, 78(4): 046110. doi: [10.1103/PhysRevE.78.046110](https://doi.org/10.1103/PhysRevE.78.046110)
39. Starnini M, Machens A, Cattuto C, Barrat A, Pastor-Satorras R. Immunization strategies for epidemic processes in time-varying contact networks[J]. *Journal of Theoretical Biology*, 2013, 337: 89–100. doi: [10.1016/j.jtbi.2013.07.004](https://doi.org/10.1016/j.jtbi.2013.07.004) PMID: [23871715](https://pubmed.ncbi.nlm.nih.gov/23871715/)
40. Ren G, Wang X. Epidemic spreading in time-varying community networks[J]. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2014, 24(2): 023116. doi: [10.1063/1.4876436](https://doi.org/10.1063/1.4876436)