Physics Letters A 381 (2017) 976–983

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Identifying multiple influential spreaders by a heuristic clustering algorithm

Zhong-Kui Bao [a], Jian-Guo Liu [b], Hai-Feng Zhang [a,c,*]

[a] School of Mathematical Science, Anhui University, Hefei 230601, PR China
[b] Data Science and Cloud Service Research Center, Shanghai University of Finance and Economics, Shanghai, 200133, PR China
[c] Department of Communication Engineering, North University of China, Taiyuan, Shan'xi 030051, PR China

ARTICLE INFO

Article history:
Received 27 September 2016
Received in revised form 12 January 2017
Accepted 26 January 2017
Available online 30 January 2017
Communicated by C.R. Doering

Keywords:
Complex networks
Multiple spreaders
Heuristic clustering algorithm

ABSTRACT

The problem of influence maximization in social networks has attracted much attention. However, traditional centrality indices are suitable for the case where a single spreader is chosen as the spreading source. Many times, spreading process is initiated by *simultaneously* choosing multiple nodes as the spreading sources. In this situation, choosing the top ranked nodes as multiple spreaders is not an optimal strategy, since the chosen nodes are not sufficiently scattered in networks. Therefore, one ideal situation for multiple spreaders case is that the spreaders themselves are not only influential but also they are dispersively distributed in networks, but it is difficult to meet the two conditions together. In this paper, we propose a heuristic clustering (HC) algorithm based on the similarity index to classify nodes into different clusters, and finally the center nodes in clusters are chosen as the multiple spreaders. HC algorithm not only ensures that the multiple spreaders are dispersively distributed in networks but also avoids the selected nodes to be very "negligible". Compared with the traditional methods, our experimental results on synthetic and real networks indicate that the performance of HC method on influence maximization is more significant.

1. Introduction

Many social, technological and biological systems can be described in terms of networks where nodes represent the elements of the systems and edges define the possible interaction patterns among nodes. The roles of nodes in social networks are often distinct, how to design effective algorithms to identify influential nodes in social networks is related to maintaining the global functionality of the system, developing efficient strategies to control epidemic spreading, accelerating information diffusion, promoting new products, and so on [1–8].

So far, lots of centrality indices have been proposed to identify influential spreaders in networks, such as degree centrality [9], betweenness centrality [10], closeness centrality [11]. Kitsak et al. proposed a k-shell decomposition method to identify the most influential spreaders, and which is better than degree centrality in many real networks [1]. But it tends to assign many nodes that have different spreading capability to the same k-shell value. In particular, which assigns all nodes of tree-like networks to 1-shell. Thus, some methods were proposed to overcome the low resolution of k-shell [6,12–14]. In addition, Radicchi et al. have shown that the nonbacktracking centrality is a highly reliable metric to identify top influential spreaders in social networks [15].

Most of the above mentioned methods mainly focus on how to find "top influential spreaders", that is to say, if one node is chosen as a *single* spreader origin, which *one* should be chosen to maximize the spreading coverage. In this case, proposed indices only need to consider the influence of node itself, but do not consider the interaction effects from other nodes. We call this situation is a single spreader case. However, it is usually that a set of different nodes are *simultaneously* chosen as spreading sources in many spreading processes, such as rumors, opinions, advertisements, and so on. Therefore, the identification of multiple influential spreaders in complex networks is also of theoretical and practical significance, however, this problem has not been well solved. An intuitive way of choosing multiple spreaders is that top ranked nodes who are sorted based on a centrality index (e.g., degree centrality, betweenness centrality, and so on) are selected. However, it may be not the optimal strategy since the top ranked nodes tend to have large overlap in their spreading process, leading to the redundancy of spreading [1,2]. For multiple spreaders case, an effective method should not only consider the influence of nodes themselves but

* Corresponding author at: School of Mathematical Science, Anhui University, Hefei 230601, PR China.
E-mail address: haifengzhang1978@gmail.com (H.-F. Zhang).

http://dx.doi.org/10.1016/j.physleta.2017.01.043
0375-9601/

also consider the dispersibility, therefore, the problem of how to identify multiple influential spreaders is more intricate. Kempe et al. have proved that the issue of finding the most multiple influential spreaders is a NP-hard optimization problem [16]. Recently, some attempts have been made along this line. For instance, Morone et al. offered a framework for the set of optimal influencers in networks by mapping the problem onto optimal percolation problem. However, the method detects the influential nodes one by one rather than *simultaneously*. Namely, the node with the highest value of CI is firstly removed, then the values of CI for remaining nodes should be recalculated, which surely increases the computation complexity [17]. Zhao et al. have proposed a method to obtain the effective multiple spreaders by generalizing the idea of the graph coloring problem to complex networks [18]. Since the distance between the multiple spreaders are not far away in sometimes, the method was improved in Ref. [19]. In the networks with community structures, Hu et al. found that the nodes with the largest degree in each community have the good performance on spreading promotion, and selecting the hub node in each community as multiple influential spreaders [20] is a good choice, but the number of communities may hinder the effectiveness of the method. Thus, how to identify the multiple influential spreaders in social networks is still an important and challenging problem [21–24].

The guiding ideology of selecting multiple spreaders is that the distance among multiple spreaders is relatively scattered, and spreaders themselves are also important. But it is almost impossible to meet both conditions together, we only try to find a tradeoff between them. In this paper, a heuristic clustering (HC) algorithm is proposed to obtain the multiple influential spreaders. In this algorithm, nodes are classified into different clusters based on one similarity index, where the number of clusters equals to the number of multiple spreaders. And the center nodes in clusters are selected as the multiple influential spreaders when the heuristic clustering process is finally stable. Experimental results in synthetic networks and real networks indicate that HC algorithm not only guarantees the selected spreaders are sufficiently scattered but also avoids to be "insignificant". Therefore, the performance of HC algorithm on influence maximization is better than other centrality indices.

The layout of the paper is as follows: Firstly, the descriptions of our method are presented in Sec. 2, and several typical centrality indices, SIR epidemic model and data sets are also introduced in this section. Then the experimental results are presented in Sec. 3. Finally, conclusions are summarized in Sec. 4.

2. Materials and methods

2.1. Heuristic clustering algorithm

An undirected and un-weighted network is represented by $G = (N, M)$ with N nodes and M edges, and its structure can be described by an adjacent matrix $A = (a_{ij})_{N\times N}$ where $a_{ij} = 1$ if node i is connected to node j, and $a_{ij} = 0$ otherwise.

If the number of multiple spreaders is m, the details of the heuristic cluster algorithm based on similarity index are the followings.

Step 1: Define similarity matrix. Because the clustering process is implemented based on the similarity between pair of nodes, a similarity matrix is defined at first. There are many ways to define the similarity, in this paper, we use the well-known local path (LP) similarity index in link prediction to define the similarity matrix, since such a similarity index provide a good tradeoff of accuracy and computational complexity [25,26]:

$$S = A^2 + \lambda \cdot A^3, \tag{1}$$

where $0 < \lambda < 1$ is a free parameter, small value of λ means less influence of long paths. $(A^2)_{xy}$ is the number of common neighbors of nodes x and y, which is also equal to the number of different paths with length 2 connecting x and y, and $(A^3)_{xy}$ is the number of different paths with length 3 connecting x and y. Hence, our HC method is a semi-local method. In our paper, we mainly set $\lambda = 0.5$, and we also check the effect of the value of λ in Figs. 5, 6 and 7.

Step 2: Form different clusters. We first randomly select m nodes as the initial centers to cluster nodes, denoted by $D = \{v_1, v_2, \cdots, v_m\}$. For each node $v_k \bar{\in} D$, the similarity $S_{v_k v_i}$ between node v_k and $v_i \in D$, $i = 1, \cdots, m$ is calculated according to Eq. (1), if there is a node $v_i \in D$ such that $S_{v_k v_i}$ is maximum, and then assign node v_k to a cluster whose center is v_i. Therefore, all nodes are classified into m clusters, denoted by $C_1, C_2, \cdots, C_m$;

Step 3: Update center of each cluster. For cluster C_t, $t = 1, \cdots, m$, according to the similarity matrix S, define the significance of node $v_x \in C_t$ in the cluster C_t as $B(x) = \sum_{v_y \in C_t} S_{v_x v_y}$, then select the node with the highest value of significance as the new center of each cluster, that is to say, the set D is updated;

Step 4: Select multiple spreaders. Repeat Step 2 and Step 3 until the algorithm is convergent. At last, the nodes in the set D are viewed as the m multiple influential spreaders.

2.2. Centrality indices

Here we briefly review the definitions of several centrality indices that will be discussed in this paper.

The degree centrality (DC) of node i is defined as the number of neighbors, namely

$$DC(i) = \sum_{j=1}^{N} a_{ij}. \tag{2}$$

The betweenness centrality (BC) of node i is defined as the fraction of all shortest paths travel through the node, which is denoted as

$$BC(i) = \sum_{s \neq i \neq l} n_{sl}^{i} / n_{sl}, \tag{3}$$

where n_{sl} and n_{sl}^{i} are the number of shortest paths between nodes s and l, and the number of shortest paths between s and l that pass through node i, respectively.

The k-shell (KS) decomposition method is implemented by the following steps: Firstly, one-degree nodes are removed and keep deleting the existing one-degree nodes until all nodes' degrees are larger than one. All of these removed nodes are 1-shell. Then remove the two-degree nodes and keep deleting until all nodes' degrees are larger than two, and include them into 2-shell. This procedure continues until all nodes have been assigned to a k-shell [1,27].

By generalizing the graph coloring in complex network, the degree coloring (DCC) method was proposed in Ref. [18,19] to identify multiple influential spreaders, which can be summarized as follows: **1)** sort the nodes in a descending order according to their degrees, such that $k(1) \geq k(2) \geq, \cdots, \geq k(N)$; **2)** define a color function π to color each node i with a color m, i.e., $\pi(i) = m$, initially, let $\pi(1) = 1$; **3)** let $C(m) = \{i | \pi(i) = m\}$, where $C(m)$ is a set containing nodes with the same color label m. If an uncolored node j is not connected to the nodes in $C(m)$, then $\pi(j) = m$; **4)** let $m := m + 1$, then choose a node at the top positions of the ranking list from the uncolored node set and back to step 3. The process ends once all the nodes are colored. Finally, the m-top

Table 1
Basic structural properties. N and M are the number of nodes and edges, respectively. $\beta_{th} = \langle k\rangle/\langle k^2\rangle$ is the epidemic/rumor threshold [28,35]. $\langle k\rangle$ is the average degree of networks. H is the degree heterogeneity, given by $\langle k^2\rangle/\langle k\rangle^2$. L and D are average shortest path length and diameter of networks, respectively.

Network	N	M	β_{th}	$\langle k\rangle$	H	L	D
SmaGri	1024	4919	0.026	9.6	3.948	2.981	6
Email	1133	5451	0.053	9.62	1.942	3.606	8
Blogs	1222	16714	0.012	27.36	2.971	2.738	8
HEP	5835	13815	0.110	4.74	1.926	7.026	19
PGP	10680	24316	0.053	4.55	4.147	7.463	24
Sex	15810	38540	0.035	4.875	5.828	5.785	17

nodes in the largest independent set are chosen as the multiple spreaders. In this paper, we also consider betweenness coloring (BCC) method and k-shell coloring method (KSC), where nodes are ranked in a descending order according to their betweenness values and k-shell values at first. That is to say, only the first step is different, and the other steps are the same as the DCC method.

2.3. Spreading models

In this paper, we use SIR epidemic model [28] to check the effectiveness of HC algorithm. For SIR epidemic model, each node can be in one of three states: susceptible, infected, or recovered. An infected node (I) can infect its *all* susceptible neighbors (S) with transmission rate β, and then becomes recovered (R) with recovery rate μ. In the classical SIR model, each infected node can contact *all* of its neighbors at per time step (we call all-contact SIR model) [28]. Here, we also consider a modified SIR model where infected nodes only contact *one* neighbor at per time step (we call single-contact SIR model) to further validate the effectiveness of our algorithm. This assumption is to mimic the situation where people can contact few of neighbors rather than all of neighbors due to their limited activity capability. In this paper, we set $\mu = 1$ for the all-contact SIR model and $\mu = 0.1$ for the single-contact SIR model.

2.4. Data description

The comparison of different algorithms are compared in six real networks, including SmaGri (citation network) [29], Email (e-mail network of University at Rovira i Virgili, URV) [30], Blogs (network of the US political blogs) [31], HEP (collaboration network of high-energy physicists) [32], PGP (an encrypted communication network) [33], Sex (the bipartite network of sex buyers and their escorts, only the giant component is chosen) [34]. For simplicity, these networks are treated as undirected and unweighed networks in this work. The detailed information about these empirical networks are presented in Table 1.

3. Results

3.1. Metric

To compare the performances of different methods in identifying multiple influential spreaders, we first select the m nodes according to one method as the spreaders (i.e., initial infected nodes). And then the epidemic spreading process is simulated according to all-contact SIR model or single-contact SIR model until there are no any infected nodes. The final recovered nodes is used to measure the performance of the method. To guarantee the reliability of the results, all of them are averaged over 100 independent realizations.

A relative ratio Δ is defined to compare the effectiveness of different methods, which is denoted as [18]

$$\Delta = \frac{R_i - R_{DC}}{R_{DC}}. \tag{4}$$

R_i is the final number of recovered nodes when multiple spreaders are chosen based on a certain method. R_{DC} is the final number of recovered nodes for the DC method. When $\Delta > 0$ means that the performance of the used method is better than the DC method, and larger value of $\Delta > 0$ indicate the better performance of the used method.

3.2. Synthetic networks

To validate the efficiency of the proposed method in identifying multiple influential spreaders, we firstly compare the HC algorithm with other methods on the two typical synthetic networks – Barabási–Alber (BA) network [36] and the Watts–Strogatz (WS) small-world networks [37], where network size is $N = 2000$ and average degree is $\langle k\rangle = 8$. The single-contact SIR model and all-contact SIR model are considered Fig. 1 and Fig. 2, respectively. Both of them indicate that the performance of HC method is better than other indices The KS index and KSC index are not considered in synthetic networks because of KS index classifying all nodes into 1-shell for BA network. All above results suggest that our HC algorithm can effectively identify multiple influential spreaders on synthetic networks.

3.3. Real networks

In the following, we implement our algorithm on the six real networks (the basic topological features are listed in Table 1) and compare our HC method with other indices for single-contact SIR model in Fig. 3. The value of Δ as a function of the transmission rates β is studied for different methods. One can find that the performance of HC method is significantly better than other methods, regardless of different number of multiples spreaders or different networks. However, the advantages of other indices are not always dominating, which may be good in some networks but lost its advantages in other networks. Moreover, we can see that the performance of KS index is generally the worst. As stated in Ref. [1], k-shell method is not a good choice in identifying multiple spreaders.

Next we study the performances of different indices for the case of all-contact SIR model. Fig. 4 indicates the performance of HC method is also better than the other indices, but the advantage is not so apparent as in single-contact SIR model. Unlike the case of single-contact SIR model where each infect node can only infect one neighbor at most at each time step, so the node's degree is not very important but the distance between multiple spreaders is the main factor. For all-contact SIR model, each infected node can contact *all* of its neighbors at per time step, thus the degree of each node is also an important factor. In the next context, we will present that our HC algorithm can ensure the distance among spreaders is very large but cannot give rise to high degree of each spreader (see Figs. 9 and 10). In addition, epidemic can quickly outbreak in networks for the all-contact SIR model, leading to the difference of among different indices is not very obvious.

3.4. Analysis of parameters

The above results are obtained by fixing the discount factor $\lambda = 0.5$, naturally, we need to check the effect of the value of λ on the effectiveness of HC method. By taking transmission rate $\beta = 0.1$ and $\beta = 0.2$ as examples, the final epidemic density σ as a function of λ for single-contact SIR model and all-contact SIR model are presented in Fig. 5 and Fig. 6, respectively. Both figures illustrate that the effect of the discount factor λ on final epidemic density is irregular. And more importantly, though some values of

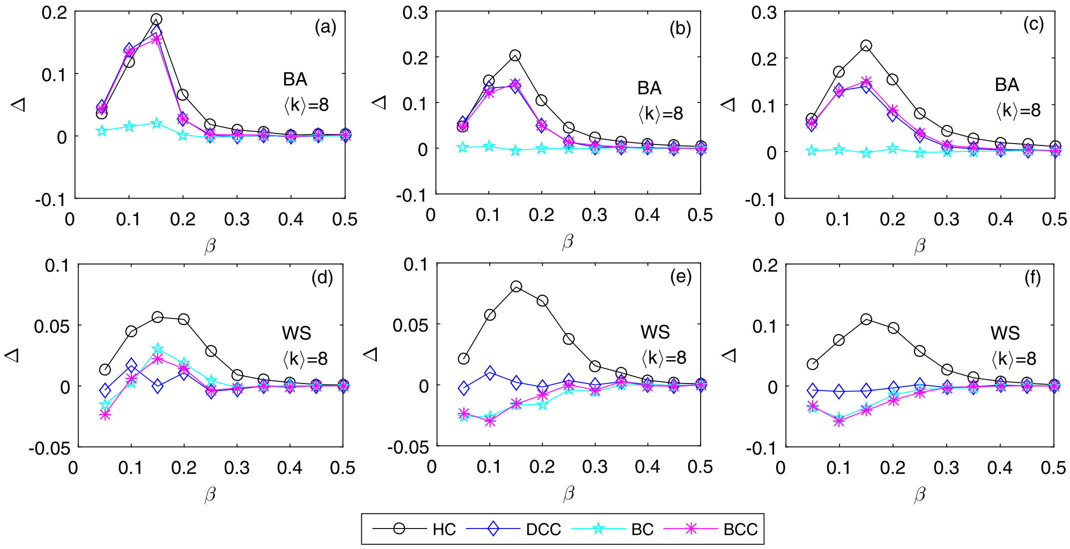

Fig. 1. (Color online) For single-contact SIR model, the relative ratios Δ for different indices as functions of transmission rate β are compared in two synthetic networks. From left panels to right panels correspond to $m = 30$, $m = 50$, $m = 90$, respectively; (a)–(c) and (d)–(f) are the results in BA network and SW network, respectively. Here recovery rate $\mu = 0.1$ and discount factor $\lambda = 0.5$. The results are averaged over 100 independent runs.

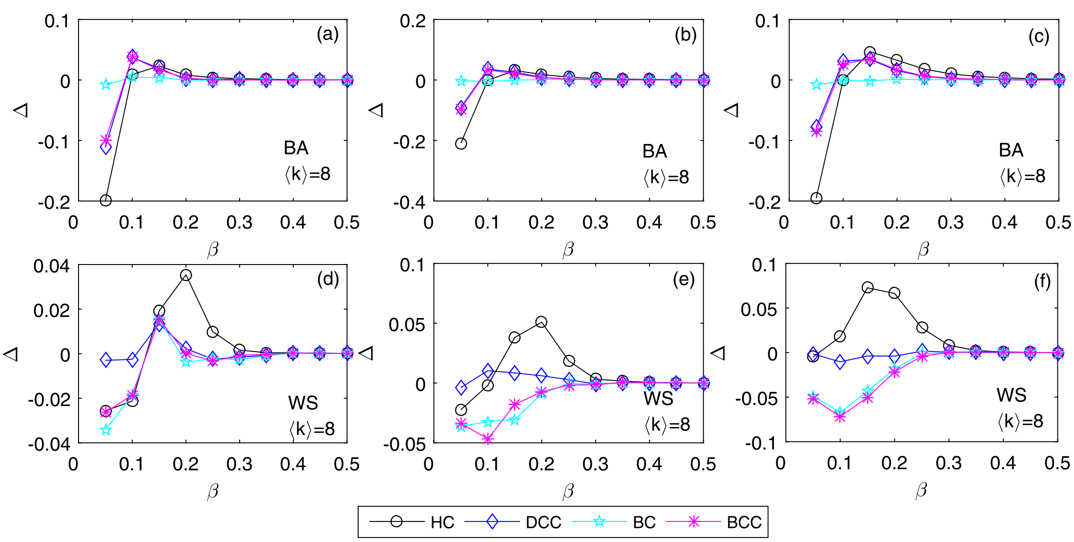

Fig. 2. (Color online) For all-contact SIR model, the relative ratios Δ regarding different indices as functions of transmission rate β are compared in two synthetic networks. From left panels to right panels to $m = 30$, $m = 50$, $m = 90$, respectively; (a)–(c) and (d)–(f) are the results in BA network and SW network, respectively. Here recovery rate $\mu = 1.0$ and discount factor $\lambda = 0.5$. The results are averaged over 100 independent runs.

λ can lead to larger value of σ, the advantage is not so remarkable. Therefore, our HC method is not very sensitive to λ. Since we cannot decide an optimal λ giving rise to the maximal value of σ in advance, we choose $\lambda = 0.5$ in above experiments is a reasonable choice.

To explain why the final epidemic density σ is not sensitively dependent on the value of λ, we investigate the effect of λ on the average distance among multiple spreaders (labeled as d_λ, right y-axis) and their average degree (labeled as $\langle k \rangle_\lambda$, left y-axis) in Fig. 7 by fixing $m = 50$. Results in Fig. 7 indicate that the average distance d_λ among multiple spreaders generally shows the reverse change trend with the average degree $\langle k \rangle_\lambda$. When multiple spreaders have higher average distance, they tend to have lower average degree. As we have mentioned in above contexts, an ideal situation for influence maximization is that the distance among multiple spreaders is large and nodes themselves are also important. However, Fig. 7 have described that changing the value of λ cannot guarantee the two conditions to meet together, which results in the effect of λ on σ is not so apparent.

We further study the effect of m on the performances of different indices in Fig. 8. As one can see that, compared with other indices, HC method can maintain significant advantage in a wide region of m, no matter of single-contact SIR model (upper panels) or all-contact SIR model (bottom panels). Moreover, the performance has an upward trend with the increasing of m in all networks, especially in SmaGri, Email, HEP, PGP and Sex networks. The results in Fig. 8 again confirm the effectiveness of HC method.

Finally, the average distance d_m and the average degree $\langle k \rangle_m$ as functions of m are illustrated in Fig. 9 and Fig. 10, respectively. On the one hand, our HC method can induce the largest value of d_m, which ensures the spreaders to be dispersively distributed in network and avoid the redundancy of spreading (see Fig. 9) [1,18]. On the other hand, although the average degree of multiple spreaders in HC method is smaller than other indices, these chosen multiple spreader are not "insignificant" nodes, because their average de-

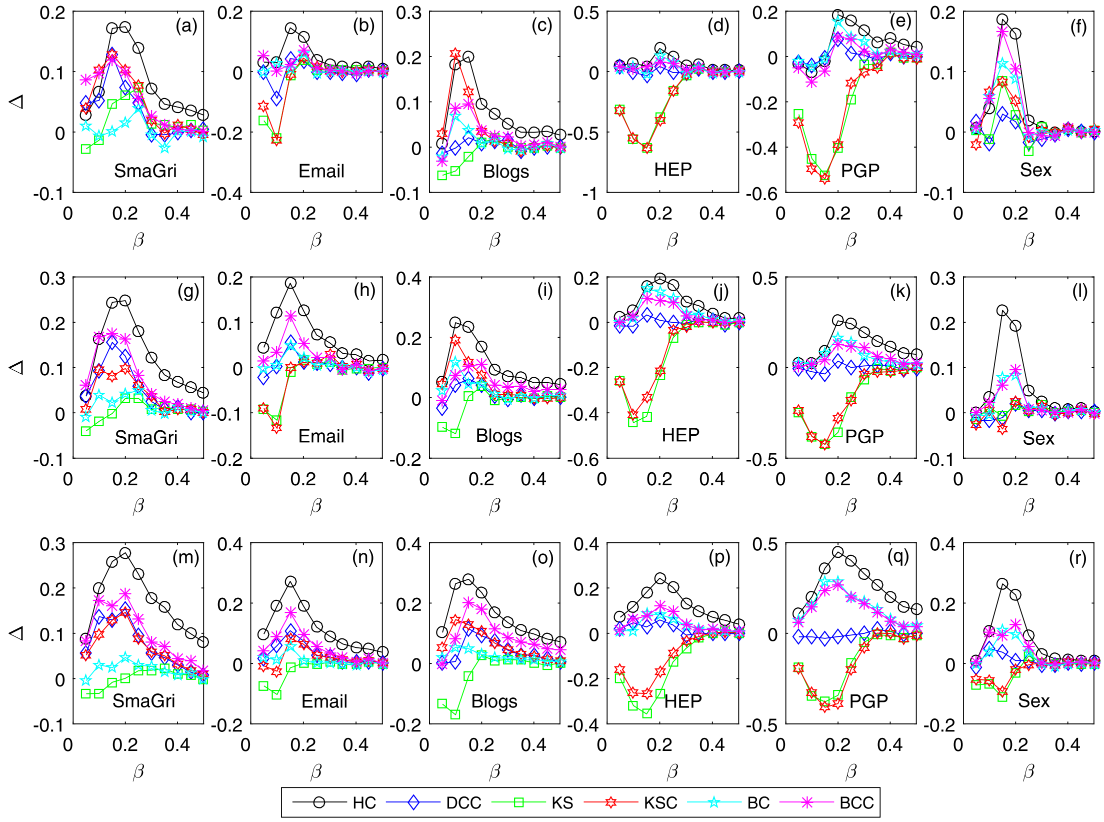

Fig. 3. (Color online) For single-contact SIR model, the relative ratios Δ as a function of transmission rate β is compared for different indices. Panels (a)–(f): the number of spreaders $m = 30$; Panels (g)–(l): the number of spreaders $m = 50$; Panels (m)–(r): the number of spreaders $m = 90$. The results are averaged over 100 independent runs. Here recovery rate $\mu = 0.1$ and discount factor $\lambda = 0.5$.

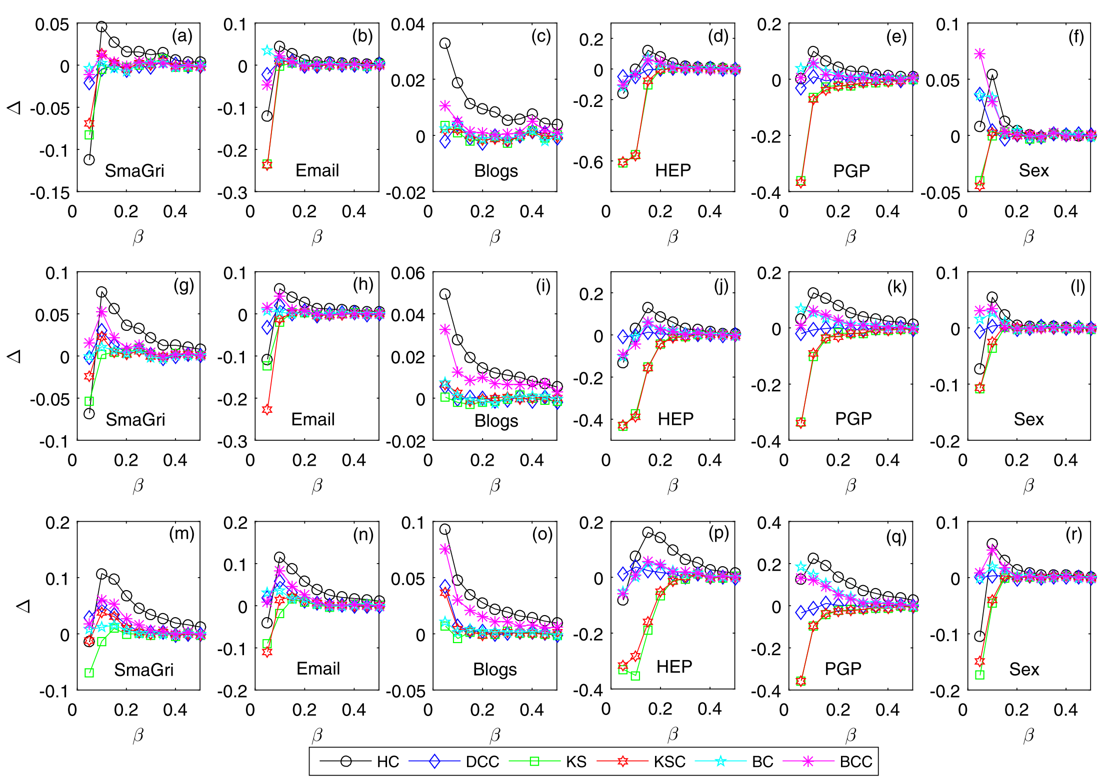

Fig. 4. (Color online) For all-contact SIR model, the relative ratios Δ for different indices as functions of transmission rate β are compared in six real networks. Panels (a)–(f): the number of spreaders $m = 30$; Panels (g)–(l): the number of spreaders $m = 50$; Panels (m)–(r): the number of spreaders $m = 90$. Here recovery rate $\mu = 1.0$ and discount factor $\lambda = 0.5$. The results are averaged over 100 independent runs.

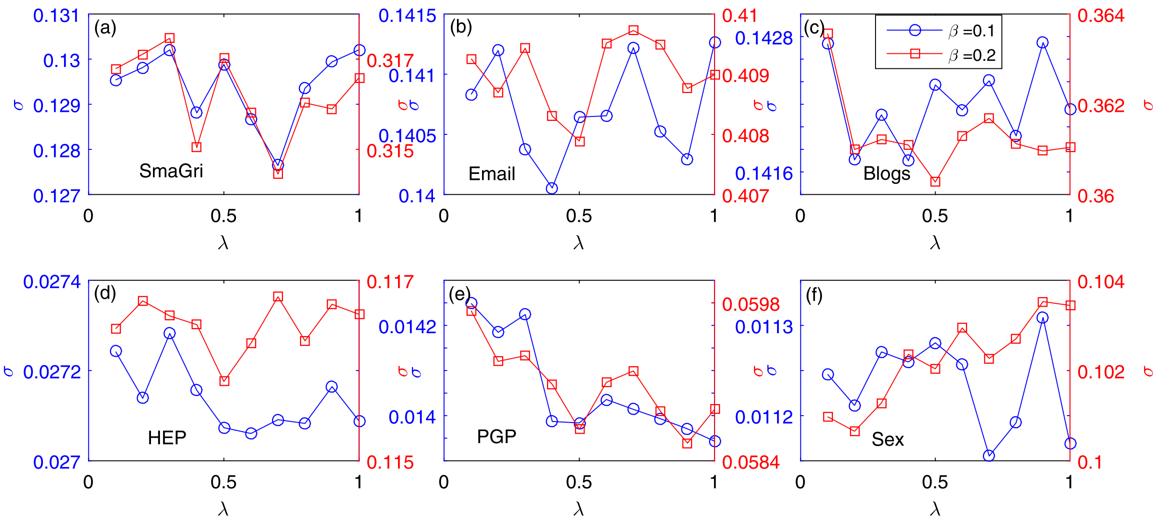

Fig. 5. (Color online) For single-contact SIR model, the final epidemic density σ as a function of λ. Left y-axis denotes the value of σ when $\beta = 0.1$, and right y-axis denotes the value of σ when $\beta = 0.2$. Here the number of spreaders is $m = 50$, and recovery rate $\mu = 0.1$. The results are averaged over 100 independent runs.

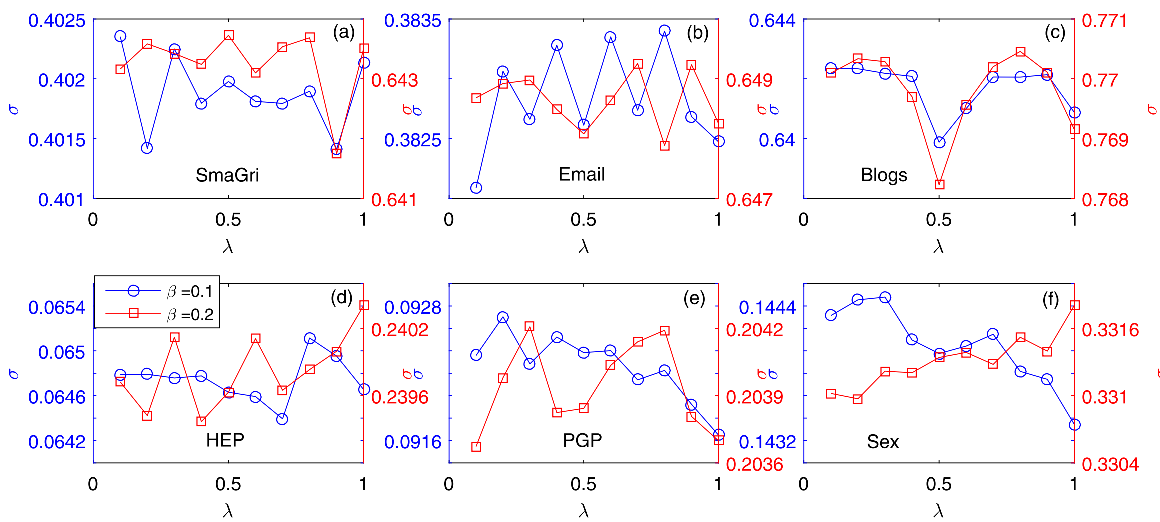

Fig. 6. (Color online) For all-contact SIR model, the final epidemic density σ as a function of λ. Left y-axis denotes the value of σ when $\beta = 0.1$, and right y-axis denotes the value of σ when $\beta = 0.2$. Here the number of spreaders is $m = 50$, and recovery rate $\mu = 1.0$. The results are averaged over 100 independent runs.

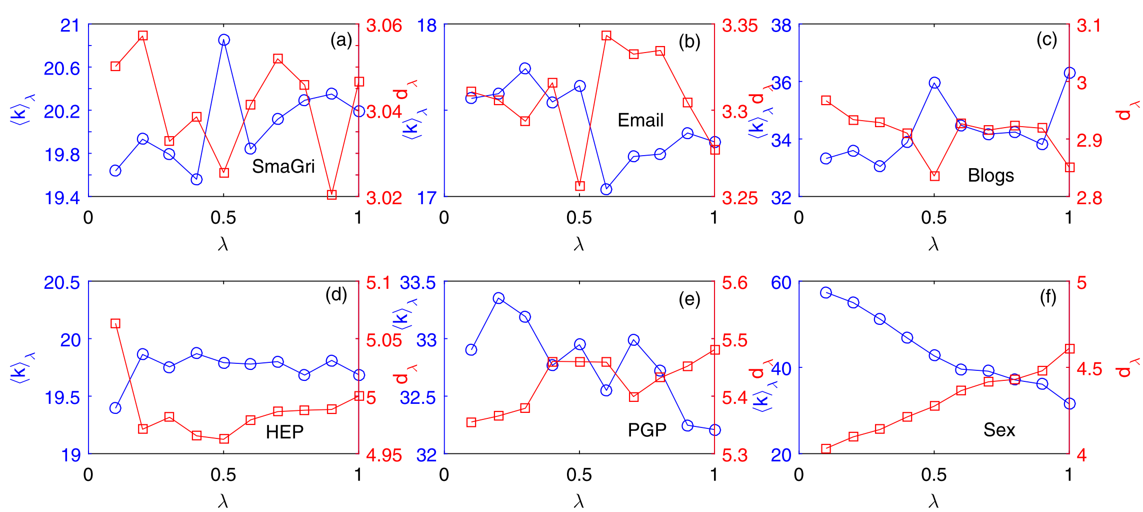

Fig. 7. (Color online) The effects of λ on the average distance d_λ and the average degree $\langle k\rangle_\lambda$ of the selected multiple spreaders are compared in six real networks. The left y-axis and the right y-axis of each subfigures denote the values of $\langle k\rangle_\lambda$ and d_λ, respectively. Here the number of spreaders is $m = 50$. The results are averaged over 100 independent runs.

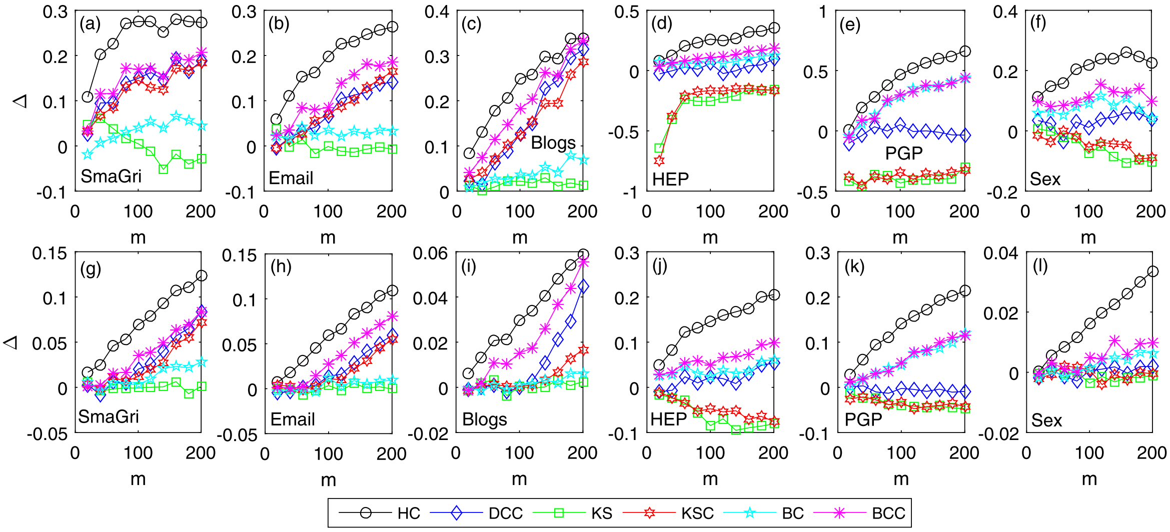

Fig. 8. (Color online) The relative ratios Δ for different indices as functions of the number of multiple spreaders m are compared in six real networks. Panels (a)–(f): single-contact SIR model; Panels (g)–(l): all-contact SIR model. Here, transmission rate $\beta = 0.2$, recovery rate $\mu = 0.1$ for single-contact SIR model and recovery rate $\mu = 1$ for all-contact SIR model. The results are averaged over 100 independent runs.

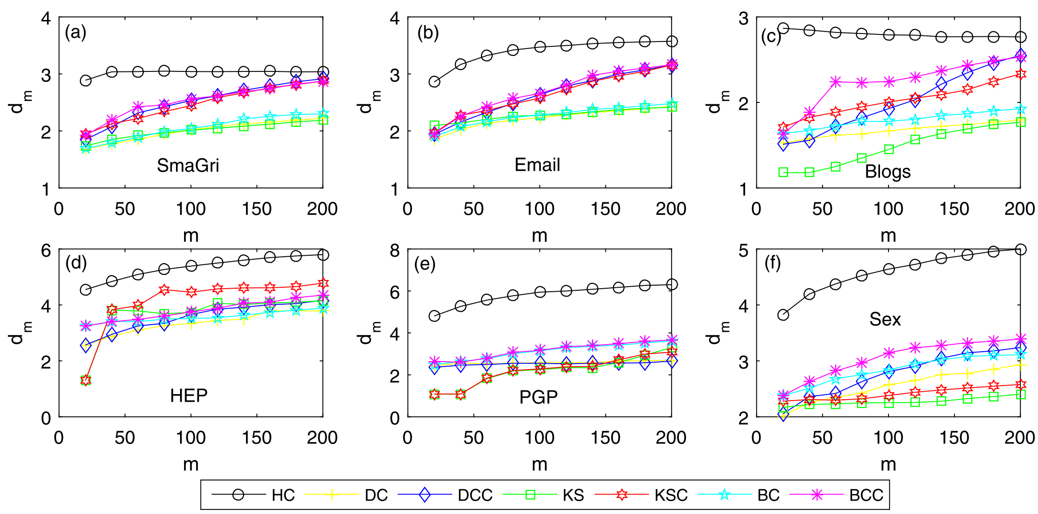

Fig. 9. (Color online) The effect of the number of multiple spreaders m on the average distance d_m is compared in six real networks. Here discount factor $\lambda = 0.5$. The results are averaged over 100 independent runs.

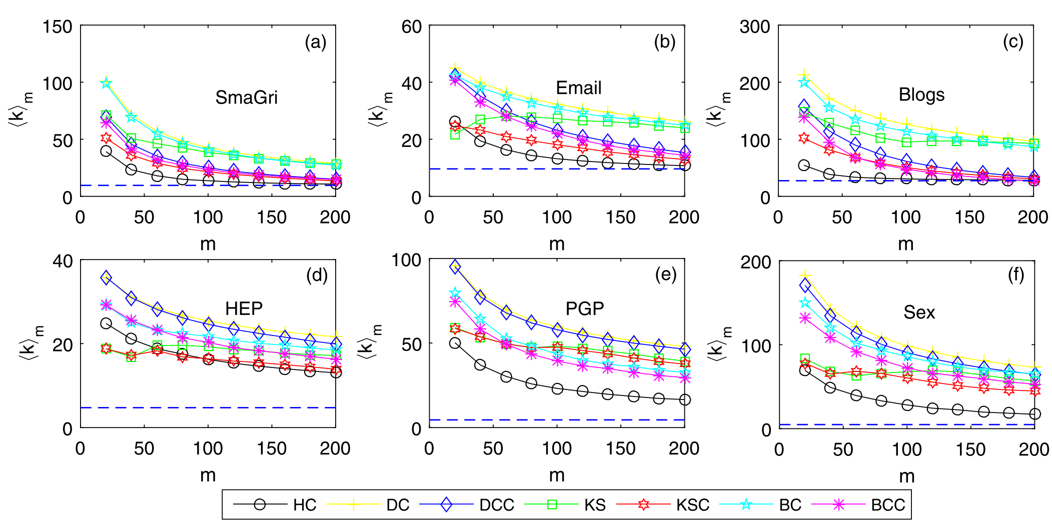

Fig. 10. (Color online) The effect of m on the average degree $\langle k \rangle_m$ is compared in six real networks. Dotted line in each subfigure denotes the average degree of the network. Here discount factor $\lambda = 0.5$. The results are averaged over 100 independent runs.

gree is still larger than the average degree of all nodes (dotted line in each subfigure), especially when the number of spreaders is not so large (see Fig. 10). Therefore, our HC method can not only ensure the dispersive distribution of the multiple spreaders but also ensure the importance of the nodes themselves to a certain extent. Our results also indicate that the average distance among spreaders may be play a more important role in multiple spreaders problem, compared with the importance of nodes themselves.

4. Conclusions

Given that how to design an effective algorithm to identify multiple influential spreaders is different from the case of single spreader, where the multiple spreaders should be scattered and the nodes themselves should be important too. However, it is hard or impossible to meet the two conditions together. In this paper, we have proposed a heuristic clustering algorithm to balance the two conditions. In the model, we first select a certain amount of nodes as initial centers, and aggregate each node into corresponding clusters by a given similarity measure. Then the center node of each cluster is updated in next step. The clustering process is repeated until the system is stable, and the centers are chosen as the multiple spreaders. Extensive experimental results in synthetic networks and real networks show that our heuristic clustering algorithm can not only make sure the spreaders are dispersively distributed in network but also ensure the importance of nodes themselves to a certain extent, leading to the best performance of HC method on influence maximization for multiple spreaders case. Therefore, our HC algorithm provides an effective way to identify the multiple influential spreaders in complex networks.

Acknowledgements

This work is funded by the NSFC (Grant Nos. 61473001, 11331009, 61374177, 71371125), the Doctoral Research Foundation of Anhui University (Grant No. 01001951) and the 211 Project of Anhui University (Grant No. 01005102).

References

[1] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse, Identification of influential spreaders in complex networks, Nat. Phys. 6 (11) (2010) 888–893.
[2] P. Basaras, D. Katsaros, L. Tassiulas, Detecting influential spreaders in complex, dynamic networks, Computer 46 (4) (2013) 24–29.
[3] L. Lü, Y.-C. Zhang, C.H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PLoS ONE 6 (6) (2011) e21202.
[4] J. Borge-Holthoefer, Y. Moreno, Absence of influential spreaders in rumor dynamics, Phys. Rev. E 85 (2) (2012) 026116.
[5] K. Klemm, M.Á. Serrano, V.M. Eguíluz, M. San Miguel, A measure of individual role in collective dynamics, Sci. Rep. 2 (2012) 292.
[6] J.-G. Liu, Z.-M. Ren, Q. Guo, Ranking the spreading influence in complex networks, Physica A 392 (18) (2013) 4154–4159.
[7] D.-B. Chen, R. Xiao, A. Zeng, Y.-C. Zhang, Path diversity improves the identification of influential spreaders, Europhys. Lett. 104 (6) (2013) 68006.
[8] Z.-M. Ren, A. Zeng, D.-B. Chen, H. Liao, J.-G. Liu, Iterative resource allocation for ranking spreaders in complex networks, Europhys. Lett. 106 (4) (2014) 48005.
[9] P. Bonacich, Factoring and weighting approaches to status scores and clique identification, J. Math. Sociol. 2 (1) (1972) 113–120.
[10] L.C. Freeman, A set of measures of centrality based on betweenness, Sociometry (1977) 35–41.
[11] G. Sabidussi, The centrality index of a graph, Psychometrika 31 (4) (1966) 581–603.
[12] A. Zeng, C.-J. Zhang, Ranking spreaders by decomposing complex networks, Phys. Lett. A 377 (14) (2013) 1031–1035.
[13] L.-L. Ma, C. Ma, H.-F. Zhang, B.-H. Wang, Identifying influential spreaders in complex networks based on gravity formula, Physica A 451 (2016) 205–212.
[14] J. Bae, S. Kim, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, Physica A 395 (2014) 549–559.
[15] F. Radicchi, C. Castellano, Leveraging percolation theory to single out influential spreaders in networks, Phys. Rev. E 93 (2016) 062314.
[16] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2003, pp. 137–146.
[17] F. Morone, H.A. Makse, Influence maximization in complex networks through optimal percolation, Nature 524 (2015) 65–68.
[18] X.-Y. Zhao, B. Huang, M. Tang, H.-F. Zhang, D.-B. Chen, Identifying effective multiple spreaders by coloring complex networks, Europhys. Lett. 108 (6) (2015) 68005.
[19] L. Guo, J.-H. Lin, Q. Guo, J.-G. Liu, Identifying multiple influential spreaders in term of the distance-based coloring, Phys. Lett. A 380 (2016) 837–842.
[20] Z.-L. Hu, Z.-M. Ren, G.-Y. Yang, J.-G. Liu, Effects of multiple spreaders in community networks, Int. J. Mod. Phys. C 25 (05) (2014) 1440013.
[21] X. Teng, S. Pei, F. Morone, H.A. Makse, Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks, preprint, arXiv:1606.02740.
[22] X. Yuan, S. Shao, H.E. Stanley, S. Havlin, How breadth of degree distribution influences network robustness: comparing localized and random attacks, Phys. Rev. E 92 (2015) 032122.
[23] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, T. Zhou, Vital nodes identification in complex networks, Phys. Rep. 650 (2016) 1–63.
[24] X. Wang, Y. Su, C. Zhao, D. Yi, Effective identification of multiple influential spreaders by degreepunishment, Physica A 461 (2016) 238–247.
[25] T. Zhou, L. Lü, Y.-C. Zhang, Predicting missing links via local information, Eur. Phys. J. B 71 (4) (2009) 623–630.
[26] L. Lü, T. Zhou, Link prediction in complex networks: a survey, Physica A 390 (6) (2011) 1150–1170.
[27] Y. Liu, M. Tang, T. Zhou, Y. Do, Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition, Sci. Rep. 5 (2015) 9602.
[28] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epidemic outbreaks in complex heterogeneous networks, Eur. Phys. J. B 26 (4) (2002) 521–529.
[29] V. Batagelj, A. Mrvar, Pajek datasets, 2006.
[30] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, A. Arenas, Self-similar community structure in a network of human interactions, Phys. Rev. E 68 (6) (2003) 065103.
[31] L.A. Adamic, N. Glance, The political blogosphere and the 2004 US election: divided they blog, in: Proceedings of the 3rd International Workshop on Link Discovery, ACM, 2005, pp. 36–43.
[32] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, N. Glance, Cost-effective outbreak detection in networks, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2007, pp. 420–429.
[33] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, A. Arenas, Models of social networks based on social distance attachment, Phys. Rev. E 70 (5) (2004) 056122.
[34] L.E.C. Rocha, F. Liljeros, P. Holme, Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts, PLoS Comput. Biol. 7 (2011) 1–9.
[35] M. Nekovee, Y. Moreno, G. Bianconi, M. Marsili, Theory of rumour spreading in complex social networks, Physica A 374 (1) (2007) 457–470.
[36] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[37] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.