Physica A 403 (2014) 130–147

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Ranking the spreading ability of nodes in complex networks based on local structure

Shuai Gao [a], Jun Ma [a], Zhumin Chen [a,*], Guanghui Wang [b], Changming Xing [a,c]

[a] School of Computer Science & Technology, Shandong University, Jinan, 250101, China
[b] School of Mathematics, Shandong University, Jinan, 250101, China
[c] School of Continuing Education, Shandong University of Finance and Economics, Jinan, 250014, China

HIGHLIGHTS

- The structure of the neighbors of a node can affect its spreading ability.
- A local structural centrality method for ranking node's spreading ability is proposed.
- The proposed method considers both the number and structure of node's neighbors.
- The proposed method outperforms other measures on both real and artificial networks.
- The proposed method is robust to different network sizes and community structure.

ARTICLE INFO

Article history:
Received 3 October 2013
Received in revised form 26 December 2013
Available online 25 February 2014

Keywords:
Complex networks
Centrality measures
Local structure
Spreading

ABSTRACT

Ranking nodes by their spreading ability in complex networks is a fundamental problem which relates to wide applications. Local metric like degree centrality is simple but less effective. Global metrics such as *betweenness* and *closeness centrality* perform well in ranking nodes, but are of high computational complexity. Recently, to rank nodes effectively and efficiently, a semi-local centrality measure has been proposed as a tradeoff between local and global metrics. However, in semi-local centrality, only the number of the nearest and the next nearest neighbors of a node is taken into account, while the topological connections among the neighbors are neglected. In this paper, we propose a local structural centrality measure which considers both the number and the topological connections of the neighbors of a node. To evaluate the performance of our method, we use the *Susceptible–Infected–Recovered* (SIR) model to simulate the epidemic spreading process on both artificial and real networks. By measuring the rank correlation between the ranked list generated by simulation results and the ones generated by centrality measures, we show that our method can rank the spreading ability of nodes more accurately than centrality measures such as degree, *k*-shell, betweenness, closeness and local centrality. Further, we show that our method can better distinguish the spreading ability of nodes.

1. Introduction

The study of the spreading process on complex networks has drawn much attention recently because of its great theoretical significance and remarkable practical value in many areas including epidemic controlling [1–5], information

* Corresponding author.
E-mail addresses: grey_shine@hotmail.com (S. Gao), chenzhumin@sdu.edu.cn (Z. Chen).

http://dx.doi.org/10.1016/j.physa.2014.02.032
0378-4371/

dissemination [6,7] and viral marketing [8,9] etc. One of the fundamental problems in understanding and controlling spreading process is evaluating the *spreading ability* for each node in the network, i.e. how many nodes will finally be covered when the spreading origins from this single node [10–16]. The knowledge of node's spreading ability shows new insights for applications such as finding social leaders [17], ranking reputation of scientists, publications [18] and designing efficient methods to either hinder epidemic spreading or accelerate information dissemination.

Over the recent years, various centrality measures such as degree, betweenness [19], closeness [20] and eigenvector [21] centralities have been proposed to rank nodes in the network. *Degree centrality* is a simple and efficient local metric, but it is less relevant since it neglects the global structure of the network. Some well-known global metrics such as *betweenness centrality* and *closeness centrality* can give better results. However due to their high computational complexity, they are incapable to be applied in large-scale networks. Recently, Kitsak et al. found that the most efficient spreaders are those located within the core of the network as identified by the *k*-shell decomposition analysis [10]. After this, some modified network decomposition algorithms have been introduced to further improve the ranking performance [16,22]. In directed networks, several iterative process based ranking methods such as PageRank [23], HITS [24] and LeaderRank [17] have been proposed to rank nodes.

Since the scale of online social systems keep growing, they can have millions or even billions of user, e.g. the total number of monthly active Facebook users is 1.1 billion till June 2013.[1] Thus the ranking algorithms which are based on global information of the network will be very time-consuming and incapable to be applied. Hence, in order to rank nodes effectively and efficiently, it is better to design the ranking algorithms based on the local information of the network. For example, a *semi-local centrality* measure which considers both the nearest and the next nearest neighbors of a node has been proposed in Ref. [11]. This centrality measure has been shown to well rank the spreading ability of nodes and achieves a good tradeoff between low-relevant degree centrality and other time-consuming measures.

However, when the local centrality is used to rank nodes, only the number of the nearest and the next nearest neighbors of a node is considered, while the topological connections among the neighbors are completely ignored. Actually, the topological connections among the neighbors are also very important. For nodes with the same local centrality, the one with denser connected neighbors is supposed to have stronger spreading ability since denser connected neighbors get more chance to influence each other. Inspired by this idea, we propose a *local structural centrality* measure which considers both the number and the topological connections of node's neighbors, where the local clustering coefficient of a node is used to measure the topological connections among its neighbors. We use the *susceptible–infected–recovered* (SIR) model [25] to simulate the epidemic spreading process on both artificial and real networks. By measuring the rank correlation between the ranked list generated by simulation results and the ones generated by centrality measures, we show that our method can rank the spreading ability of nodes more accurately than centrality measures such as degree, *k*-shell, betweenness, closeness and local centrality. Through experiments on artificial networks generated by Barabási–Albert (BA) [26,5] network model and Lancichinetti–Fortunato–Radicchi (LFR) network model [27], we show that our method can outperform other centrality measures in scale-free networks with different sizes and different community structure. Further, we show that our proposed method can better rank the most influential nodes than other measures considered. Moreover, we use the *susceptible–infected* (SI) model [25] to simulate the epidemic spreading process and show that our proposed method can better rank the spreading ability of nodes under the SI model. Finally, we examine the ability of different methods to distinguish the spreading ability of the nodes and show that our proposed method performs better.

Following parts are organized as follows. We briefly review the definition of centrality measures used for comparison in Section 2 and introduce our local structural centrality measure in Section 3. In Section 4, we present the data, the spreading model and the evaluation measure that are used to evaluate the performance of our method. The experimental results are presented in Section 5. We conclude our paper and give a discussion in Section 6.

2. Centrality measures

Consider an unweighted and undirected simple network $G = (V, E)$ with $n = |V|$ nodes and $m = |E|$ links. G could be described by an adjacent matrix $A = \{a_{uv}\} \in R^{n,n}$, where $a_{uv} = 1$ if node u is connected with node v and $a_{uv} = 0$ otherwise. We use $\Gamma_h(v)$ to denote the set of neighbors within h-hops from node v.

The *degree centrality* (DC), $C_D(v)$, of node v can be calculated as

$$C_D(v) = \sum_{u=1}^{n} a_{uv} = |\Gamma_1(v)|. \tag{1}$$

The computational complexity for degree centrality is $O(n)$.

The *k-shell centrality* (KS) [10], C_{KS}, is obtained by employing *k*-shell decomposition algorithm on the network. The algorithm can be performed iteratively. First, we remove all nodes with degree equals to 1 and assign their $C_{KS} = 1$. After removing all these nodes, some nodes may be left with one link, so we keep pruning the network until there is no node with degree equals to 1 left in the network. The removed nodes, together with their corresponding links, form a *k* shell with

[1] http://www.statisticbrain.com/facebook-statistics/.

$C_{KS} = 1$. In a similar fashion, we iteratively remove the next k shell, $C_{KS} = 2$, and continue removing higher k shells until all nodes are removed. As a result, each node is associated with one C_{KS} and the network can be viewed as the union of all k shells. The computational complexity for k-shell centrality is $O(m)$ [28].

The *betweenness centrality* (BC) [19], $C_B(v)$, of node v is defined as the fraction of shortest paths between all node pairs that pass through node v which is given by

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

where σ_{st} denotes the number of all possible shortest paths between node s and t, and $\sigma_{st}(v)$ denotes the number of shortest paths between node s and t which pass through node v. For unweighted networks, calculating betweenness centrality costs $O(nm) = O(n^2\langle k\rangle)$ using Brandes' algorithm [29], where $\langle k\rangle$ is the average degree of the network.

The *closeness centrality* (CC) [20], $C_C(v)$, of node v is defined as the reciprocal of the sum of the shortest distances to all other nodes of V.

$$C_C(v) = \frac{1}{\sum_{u \in V \setminus v} d_{uv}} \tag{3}$$

where d_{uv} denotes the shortest distance between node u and v. The calculation of closeness centrality takes the complexity $O(n^3)$ with the Floyd's algorithm [30]. It can be more efficiently calculated in a sparse graph using Johnson's algorithm [31], which takes $O(n^2 \log n + nm)$.

The degree centrality is a simple and efficient local metric, but it is less relevant since it neglects the global structure of the network. Both betweenness and closeness centrality consider the global information and can give better ranking results. But they are of high computational complexity since they need to calculate the shortest paths between all pairs of nodes in the network which is very time-consuming. As a tradeoff, a semi-local centrality measure by considering both the nearest and the next nearest neighbors is proposed in Ref. [11]. The *local centrality* (LC), $C_L(v)$, of node v is defined as

$$C_L(v) = \sum_{u \in \Gamma_1(v)} \sum_{w \in \Gamma_1(u)} N(w) \tag{4}$$

where $N(w)$ is the number of the nearest and the next nearest neighbors of node w, i.e. $N(w) = |\Gamma_2(w)|$. The calculation of $N(w)$ needs traversing node w's neighborhood within two steps, so the total computational complexity for local centrality is $O(n\langle k\rangle^2)$ [11].

3. Our proposed local structural centrality

To begin our analysis, we first discuss the impact of the local structure of a node on its spreading ability. By local structure, we mean the structure of the nearest and next nearest neighbors of a node. As is known to us, the number of neighbors of a node can affect its spreading ability. For example, the degree centrality considers the number of the nearest neighbors and the local centrality considers the number of the nearest and the next nearest neighbors. Both of them are commonly used centrality measures. Besides the number of neighbors, we further investigate the impact of the topological connections among neighbors on node's spreading ability. For better illustration, we give two toy networks in Fig. 1. Considering node i and node j in Fig. 1(a), they have the same number of neighbors while some neighbors of node j are connected with each other. In epidemic spreading, once node i and node j are infected, the neighbors of node i can only be infected by node i, while the neighbors of node j can be infected by other connected infected neighbors once node j fails to infect them. Intuitively, node j has stronger spreading ability than node i. Considering a larger network in Fig. 1(b), we analyze the spreading ability of node 1 and node 8. They have the same degree ($C_D(1) = C_D(8) = 6$) and the same local centrality ($C_L(1) = C_L(8) = 130$), nonetheless node 8 is expected to have stronger spreading ability since two of its neighbors, node 13 and node 14, have denser connected neighbors than node 6 and node 7, which are neighbors of node 1 on the symmetry position. The simulation results using SIR model can be found in Fig. 1(b) which confirm the above expectation.

Seeing that the topological connections among neighbors of a node have an impact on its spreading ability, we propose a new centrality measure, called *local structural centrality* (LSC), which considers both the number of nearest and the next nearest neighbors, and the topological connections among neighbors. The new centrality measure is defined as

$$\begin{aligned} C_{LS}(v) &= \sum_{u \in \Gamma_1(v)} Q(u) \\ &= \sum_{u \in \Gamma_1(v)} \left(\alpha N(u) + (1-\alpha) \sum_{w \in \Gamma_2(u)} c_w \right) \end{aligned} \tag{5}$$

where $N(u) = |\Gamma_2(u)|$ is the number of nearest and next nearest neighbors of node u, c_w represents the local clustering coefficient of node w and α is a tunable balance parameter between 0 and 1. The local clustering coefficient of node w

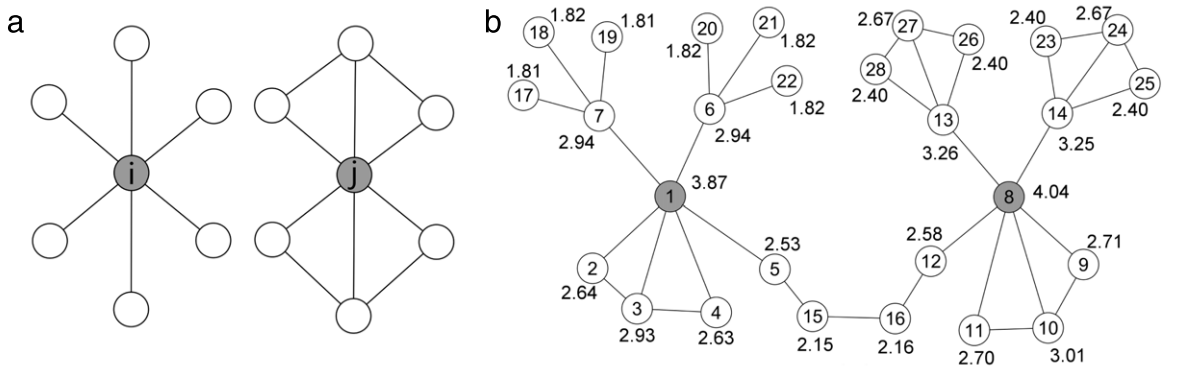

Fig. 1. The example of toy networks. (a) Node i and node j have the same degree ($C_D(i) = C_D(j) = 6$) while the spreading ability of node i and node j are 2.72 and 3.22 respectively. (b) Node 1 and node 8 have the same degree ($C_D(1) = C_D(8) = 6$) and same local centrality ($C_L(1) = C_L(8) = 130$) but the different local structural centrality ($C_{LS}(1) = 35.04$, $C_{LS}(8) = 37.64$, with $\alpha = 0.7$), while the spreading ability of node 1 and node 8 are 3.87 and 4.04 respectively. Spreading ability presented around each node is estimated at $\beta = \beta^c_{rand}$ (0.28 in (a) and 0.31 in (b)) by performing 10^6 independent runs of the SIR model per seed node, where β^c_{rand} is the epidemic threshold.

quantifies how close its neighbors are to being a clique (complete graph) [32], which is defined as

$$c_w = \frac{2|\{e_{ij} : i, j \in \Gamma_1(w), e_{ij} \in E\}|}{C_D(w)(C_D(w) - 1)}. \tag{6}$$

Here we use the local clustering coefficient of a node to evaluate the topological connections among its nearest neighbors.

To gain a better understanding of our new centrality measure, we take a further look into Eq. (5). The local structural centrality value of node v is denoted by $C_{LS}(v)$. For each neighbor node u of node v, we take $Q(u)$ as node u's contribution to the final centrality value of node v. To calculate $Q(u)$ for each node u, we consider its nearest and next nearest neighbor set $\Gamma_2(u)$. For each node $w \in \Gamma_2(u)$, we split its contribution to $Q(u)$ into two parts: direct contribution and indirect contribution. As we mentioned before, the new centrality measure considers both the number of the neighbor nodes and the topological connections among the neighbor nodes. Here, the direct contribution corresponds to the number of neighbor nodes and the indirect contribution corresponds to the topological connections among the neighbor nodes. Intuitively, the direct contribution measures how many neighbor nodes the initially infected node can infect and the indirect contribution measures to what extent the neighbor nodes can infect each other. Specifically, the direct contribution of node w to $Q(u)$ is simply 1, namely each node in $\Gamma_2(u)$ is equally treated and only counted once in the calculation of $Q(u)$. The indirect contribution of node w to $Q(u)$ is its local clustering coefficient c_w. Note that the local clustering coefficient of node w measures how close its neighbors are connected to each other, it can be well applied to measure the indirect influence among neighbors. We set an global parameter α to balance the direct and indirect contribution. The total contribution of node w to $Q(u)$ is $\alpha \cdot 1 + (1 - \alpha) \cdot c_w$. By summing contributions of all the nodes in $\Gamma_2(u)$ together, we get

$$\begin{aligned} Q(u) &= \underbrace{\alpha \cdot 1 + (1-\alpha) \cdot c_{w_1}}_{\text{contribution of } w_1} + \underbrace{\alpha \cdot 1 + (1-\alpha) \cdot c_{w_2}}_{\text{contribution of } w_2} + \cdots + \underbrace{\alpha \cdot 1 + (1-\alpha) \cdot c_{w_{N(u)}}}_{\text{contribution of } w_{N(u)}} \\ &= \alpha N(u) + (1-\alpha) \sum_{w \in \Gamma_2(u)} c_w \end{aligned} \tag{7}$$

where $w_1, w_2, \ldots, w_{N(u)} \in \Gamma_2(u)$. Then, by summing all the $Q(u)$ for each neighbor node u of node v, we get the final form of local structural centrality $C_{LS}(v)$ as defined in Eq. (5).

Again, we take Fig. 1(b) as an example and focus on node 1 and node 8. By applying our local structural centrality measure with $\alpha = 0.7$, we can get $C_{LS}(1) = 35.04$ and $C_{LS}(8) = 37.64$. Note that the spreading ability of node 1 and node 8 are 3.91 and 4.05 respectively, our proposed method can well discriminate these two nodes and rank them correctly. Further, we show the values of all six centrality measures for the top-10 nodes with the strongest spreading ability in Table 1. The spreading ability s^β of each node is simulated using SIR model with the spreading probability $\beta = \beta^c_{rand} = 0.31$, where β^c_{rand} is the epidemic threshold of the network. We can see that the proposed local structural centrality measure can better rank the spreading ability of nodes than other centrality measures considered.

Before talking about the performance of our method, we first show its efficiency by investigating its computational complexity. The calculation of $N(u)$ needs traversing node u's neighborhood within two steps, and it costs $O(n\langle k\rangle^2)$. The computational complexity for calculating the local clustering coefficient for a single node is $O(\langle k\rangle^2)$, where $\langle k\rangle$ is the average degree of the network. So the total computational complexity for our centrality measure is $O(n\langle k\rangle^2)$. Our method takes the same complexity as the local centrality and costs much lower complexity than betweenness and closeness centrality. Further, we show the execution time for all centrality measures on four real networks in Table 2. We can see that our LSC method costs slightly more time than LC method and far less time than KS, BC and CC method especially when the size of network becomes larger. Of course, the execution time for DC is the shortest.

Table 1
The top-10 nodes with the strongest spreading ability s^{β} and their corresponding centrality values by degree centrality (DC), k-shell method (KS), betweenness centrality (BC), closeness centrality (CC), local centrality (LC) and our local structural centrality (LSC with $\alpha = 0.7$). s^{β} is estimated at $\beta = \beta^{c}_{rand} = 0.31$ by performing 10^6 independent runs of the SIR model per seed node, where β^{c}_{rand} is the epidemic threshold.

Node	s^{β}	DC	KS	BC	CC	LC	LSC
8	4.04	6	2	433	0.0096	130	37.64
1	3.87	6	2	433	0.0096	130	35.04
13	3.26	4	2	145	0.0081	87	22.12
14	3.25	4	2	145	0.0081	87	22.12
10	3.01	3	2	1	0.0078	82	21.58
7	2.94	4	1	150	0.0081	71	18.42
6	2.94	4	1	150	0.0081	71	18.42
3	2.93	3	2	1	0.0078	82	19.38
9	2.71	2	2	0	0.0078	69	16.74
11	2.70	2	2	0	0.0078	69	16.74

Table 2
The execution time (in seconds) for all centrality measures on all four real networks.

Network	DC	KS	BC	CC	LC	LSC
Email	0.03	1.61	3.18	3.24	0.76	1.34
Blog	0.04	4.60	10.20	10.62	1.03	1.73
PGP	0.06	37.57	113.13	118.10	3.75	6.02
Twitter	0.09	3085.71	12324.16	13980.98	68.38	79.40

Table 3
The basic topological features of the giant connected components of the four real networks. Structural properties include node number (n), edge number (m), average degree ($\langle k\rangle$), max degree ($k_{\max}$), clustering coefficient (C) [32], average shortest path length ($\langle d\rangle$), assortative coefficient (r) [37], degree heterogeneity ($H = \frac{\langle k^2\rangle}{\langle k\rangle^2}$) [38] and epidemic threshold ($\beta^{c}_{rand} = \frac{\langle k\rangle}{\langle k^2\rangle}$) [39].

Network	n	m	$\langle k\rangle$	$k_{\max}$	C	$\langle d\rangle$	r	H	β^{c}_{rand}
Email	1 133	5 451	9.62	71	0.2202	3.6028	0.0782	1.9421	0.0535
Blog	3 982	6 803	3.42	189	0.2838	6.2504	−0.1330	4.0381	0.0725
PGP	10 680	24 316	4.55	205	0.2659	7.4848	0.2382	4.1465	0.0530
Twitter	65 911	185 737	5.64	233	0.1823	7.8136	0.3157	2.5829	0.0687

4. Experimental setup

To evaluate the performance of our proposed centrality measure, we apply it on both artificial and real networks. The artificial networks include networks generated by the Barabási–Albert (BA) network model [26,5] and the Lancichinetti–Fortunato–Radicchi (LFR) network model [27]. All these artificial networks are undirected unweighted and scale-free. The real networks are drawn from disparate fields, including: (i) Email: a network of e-mail interchanges between members of the University Rovira i Virgili (Tarragona) [33]. (ii) Blog: a network of the communication relationships between owners of blogs on the MSN (Windows Live™) Spaces website [34]. (iii) PGP: an encrypted communication network. Pretty-Good-Privacy algorithms have been developed in order to maintain privacy between peers, wherefore, it is also called web of trust of PGP [35]. (iv) Twitter: the following relationships among users on Twitter.com, a widely used microblogging system [36]. We treat all these real networks as undirected and only consider their giant connected components. The basic topological features of the giant connected component of these networks are summarized in Table 3. Brief definitions of the monitored topological measures can be found in the table caption.

The ranked lists are obtained by applying all centrality measures on each network. In principle, the ranked list generated by an effective ranking method should be as consistent as possible with the ranked list generated by the real spreading process. In order to simulate the real spreading process, we use the standard SIR model [25]. In the SIR model, all nodes are initially in the *susceptible* state (S) except for one node in the *infected* state (I). At each time step, the infected nodes infect their susceptible neighbors with probability β and then enter the *recovered* state (R) with probability equal to 1, where they become immunized and cannot be infected again [10]. The spreading process ends when there is no infected node existed in this network. The spreading ability of node v, s^{β}_{v}, is defined as the number of nodes that are finally infected at the end of spreading process which originates from node v. Then, we can obtain a ranked list by ranking the spreading ability of nodes in the network. Further, we use $\bar{s_v}$ to denote the average spreading ability of node v on the entire range of β.

In order to evaluate how the ranked list generated by a certain centrality measure is correlated to the ranked list generated by the real spreading ability s^{β} of the nodes, we use the Kendall's tau coefficient (τ) [40]. The Kendall's tau coefficient considers a set of joint observations from two random variables X and Y (in our case, X can be the values of a certain centrality measure and Y can be simulation results for all nodes). Any pair of observations (x_i, y_i) and (x_j, y_j) are said to be

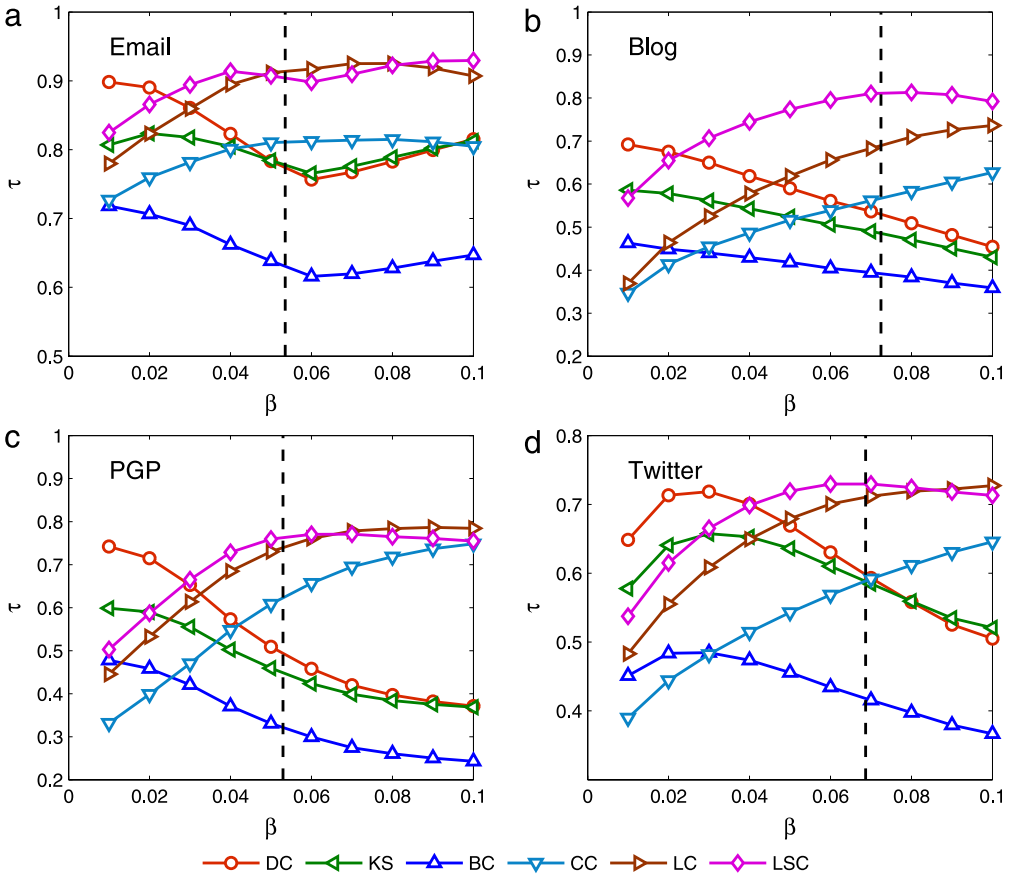

Fig. 2. (Color online) The Kendall's tau τ values obtained by comparing the ranked lists generated by all six centrality measures and the ranked list generated by the SIR spreading process on four real networks, where the vertical dashed line corresponds to the epidemic threshold β^c_{rand}. Centrality measures are degree centrality (DC), k-shell centrality (KS), betweenness centrality (BC), closeness centrality (CC), local centrality (LC) and our local structural centrality (LSC with $\alpha = 0.2$). The results are obtained by averaging over 1000 independent realizations where the spreading probability β is ranging from 0.01 to 0.1.

Table 4
The $\langle\tau\rangle$ values for all six centrality measures on all four real networks.

Network	$\langle\tau\rangle_{DC}$	$\langle\tau\rangle_{KS}$	$\langle\tau\rangle_{BC}$	$\langle\tau\rangle_{CC}$	$\langle\tau\rangle_{LC}$	$\langle\tau\rangle_{LSC}$
Email	0.8177	0.7982	0.6563	0.7937	0.8862	**0.8994**
Blog	0.5767	0.5139	0.4110	0.5134	0.6065	**0.7466**
PGP	0.5218	0.4655	0.3386	0.5911	0.6900	**0.7064**
Twitter	0.6262	0.5974	0.4342	0.5422	0.6556	**0.6850**

concordant if the ranks for both elements agree: that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. They are said to be discordant if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant. The Kendall's tau coefficient τ is defined as:

$$\tau = \frac{n_c - n_d}{0.5n(n-1)} \tag{8}$$

where n_c and n_d denote the number of concordant and discordant pairs respectively. Higher τ value indicates more accurate ranked list a centrality measure could generate. The most ideal case is $\tau = 1$, where the ranked list generated by the centrality measure is exactly the same as the ranked list generated by the real spreading process.

5. Results

In this paper, we use relatively small values of β in SIR model, namely $\beta \in (0, 0.1]$, so that the infected percentage of the nodes remains small. When $\beta = 0.1$, the average infected percentage of the nodes is 0.1147 in Email, 0.0008 in Blog, 0.0057 in PGP and 0.0031 in Twitter. In the case of large β values, where spreading can reach a large fraction of the nodes, the

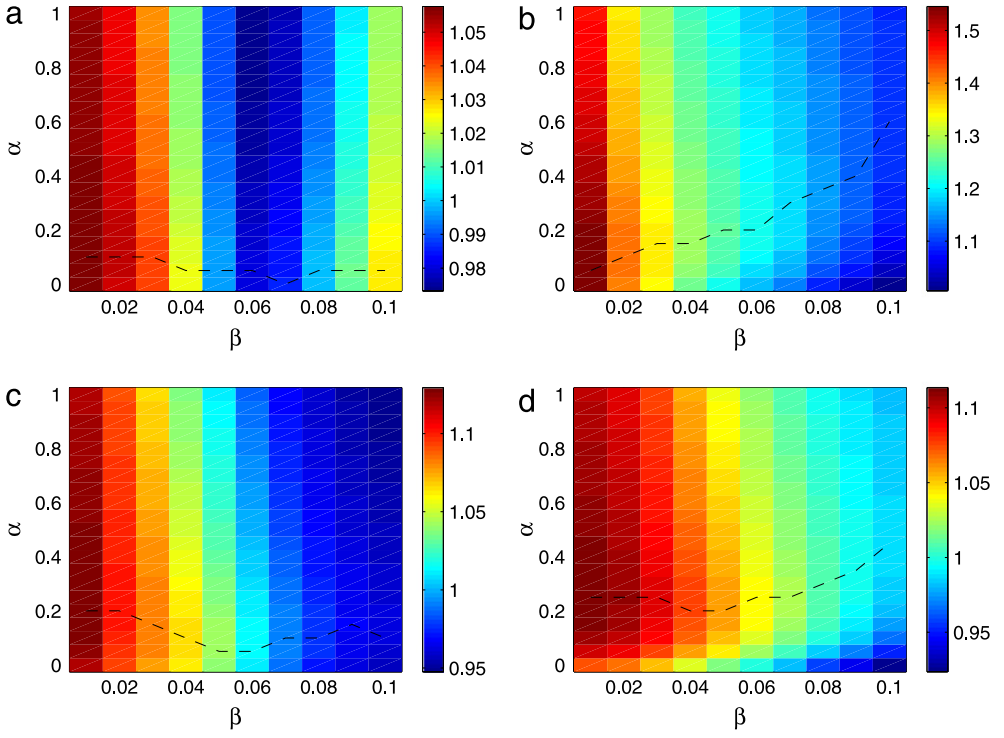

Fig. 3. (Color online) The value of τ' under different parameters α and β. The networks are (a) Email, (b) Blog, (c) PGP and (d) Twitter. The black dashed lines mark the optimal α under different β. The results are obtained by averaging over 1000 independent realizations.

role of individual node is no longer important and spreading would cover almost all the network, independently of where it originates from Ref. [10]. Specially, we are more interested in the area when β is around the epidemic threshold β^c_{rand} since in this area the initially infected nodes are perturbations that can trigger spreading at all scales [41]. In the following subsections, we first fix the balance parameter α in our LSC method to be 0.2 and compare its performance with other centrality measures, and then systematically study how the parameter α affects the performance of the LSC method in Section 5.1. In Section 5.2, considering sometimes people are only interested in the most influential nodes in the network, we show the performance of our method in ranking the most influential nodes. Moreover, we show the performance of our method on artificial networks generated by BA and LFR network model in Section 5.3, to investigate the impact of the size and the community structure of the network on the performance of our method. Besides, we check the performance of our proposed method on the standard SI model in Section 5.4. Further, we show the ability of our method in distinguishing the spreading ability of nodes in Section 5.5.

5.1. Effectiveness

By fixing the balance parameter α to be 0.2 in our LSC method, we compare its τ values with the τ values for DC, KS, BC, CC and LC under different spreading probabilities β and show the results in Fig. 2. We can see that our local structural centrality can achieve the best performance on a wide range of spreading probability β in four real networks, especially when β is around the epidemic threshold β^c_{rand}. When β is far smaller than β^c_{rand}, the spreading is typically confined to the neighborhood of the initially infected node, hence the nodes with larger degree can infect more nodes, that explains why degree centrality always achieve the largest τ value when β is too small (less than 0.02). When β grows larger, our method begins to show better performance. As is shown in Fig. 2, our proposed method can achieve the best performance on a wide range of β in Email and Blog networks. In PGP and Twitter networks, our method can achieve the best performance when β is around epidemic threshold β^c_{rand}, and when β grows larger than β^c_{rand}, our method becomes less effective than local centrality. We argue that, when β is larger than β^c_{rand}, the epidemics can spread to much farther away from the initially infected node. The local centrality measure which considers nodes within 4-hops away from the initially infected node contains more information of neighbor nodes than our method. That explains why our method is less effective than local centrality under this circumstance. Even though, since our method considers local structural information, it can still achieve the best performance on a wide range of β.

Since different real networks may ask for different α for the LSC method to achieve its best performance, we vary the value of α from 0 to 1 to study how it affects the performance of our LSC method. We calculate the ratio of τ_{LSC} and τ_{LC} as $\tau' = \tau_{LSC}/\tau_{LC}$ to show the advantage of our method more clearly. The τ' values under various α and β are shown in

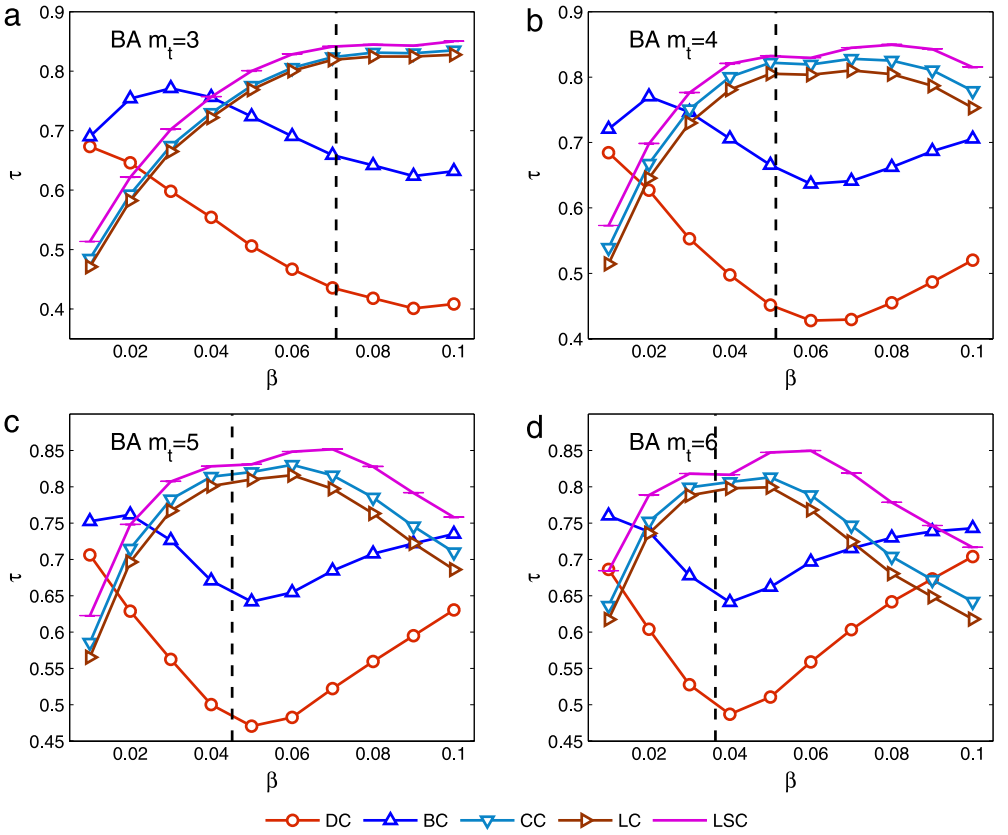

Fig. 4. (Color online) The Kendall's tau τ values for five centrality measures on four BA networks with $n = 5000$ and $m_t = 3, 4, 5, 6$, where the vertical dashed line corresponds to the epidemic threshold β^c_{rand}. Centrality measures are degree centrality (DC), betweenness centrality (BC), closeness centrality (CC), local centrality (LC) and our local structural centrality (LSC). We show the performance of our LSC method with α varying from 0 to 1 using box plots. Boxes cover 50% of data and whisker cover 95%. The line in a box represents the median of data. One can observe that our LSC method shows the best performance on a wide range of β on all four networks. Since the clustering coefficients for these BA networks are too small (0.006–0.009), even α varies, the final centrality value of the target node does not change too much, hence the performance of our method slightly changes with the varying of α. The results are obtained by averaging over 1000 independent realizations where the spreading probability β is ranging from 0.01 to 0.1.

Fig. 3. One can immediately notice that the ratio τ' can be larger than 1 on a wide range of β, which further confirms the results shown in Fig. 2. Also, we can observe that, given a certain spreading probability β, there is always an optimal α which can achieve the highest τ' value. The black dashed curves in Fig. 3 show the change of optimal α with different β on four networks. It is worth noting that most of the optimal α are smaller than 0.5. Since α serves as an balance parameter in the definition of the proposed LSC method in Eq. (5), when α takes a small value, the proposed method assigns more weight on the local clustering coefficient part, which confirms our intuition that the local structural information can play an important role in measuring the spreading ability of a node.

To reflect the overall spreading ability of all nodes, we calculate $\langle\tau\rangle$ by averaging the τ values over the entire range of spreading probability β we considered. We report the results of this measure on all four real networks in Table 4. As we can see, our local structural centrality measure can achieve the best $\langle\tau\rangle$ values on all four networks, which certifies its effectiveness in ranking the spreading ability of nodes in the network.

5.2. Rank the most influential nodes

The calculation of $\langle\tau\rangle$ value takes all the nodes in the network into account. Sometimes, people are only interested in the most influential nodes in the network [10]. Therefore, we investigate another measurement $\langle\hat{\tau}\rangle$, which only considers the most influential nodes (i.e. nodes with strongest average spreading ability $\bar{s}$ which is estimated by averaging the spreading ability s^β on entire range of β using the SIR-model-based simulation). In $\langle\hat{\tau}\rangle$, we only consider L (from 10 to 1000) nodes with the strongest average spreading ability in the network. Except that, the calculation of $\langle\hat{\tau}\rangle$ is exactly the same as $\langle\tau\rangle$. The results on four networks are shown in Fig. 8. We can see that, our proposed method can achieve the best $\langle\hat{\tau}\rangle$ value on almost entire range of L in Email, Blog and PGP networks. In Twitter, our proposed measure shows almost the same performance as the local centrality measure and performs slightly better when L is smaller than 800. Generally speaking, our local structural centrality measure can better rank the most influential nodes than other measures considered.

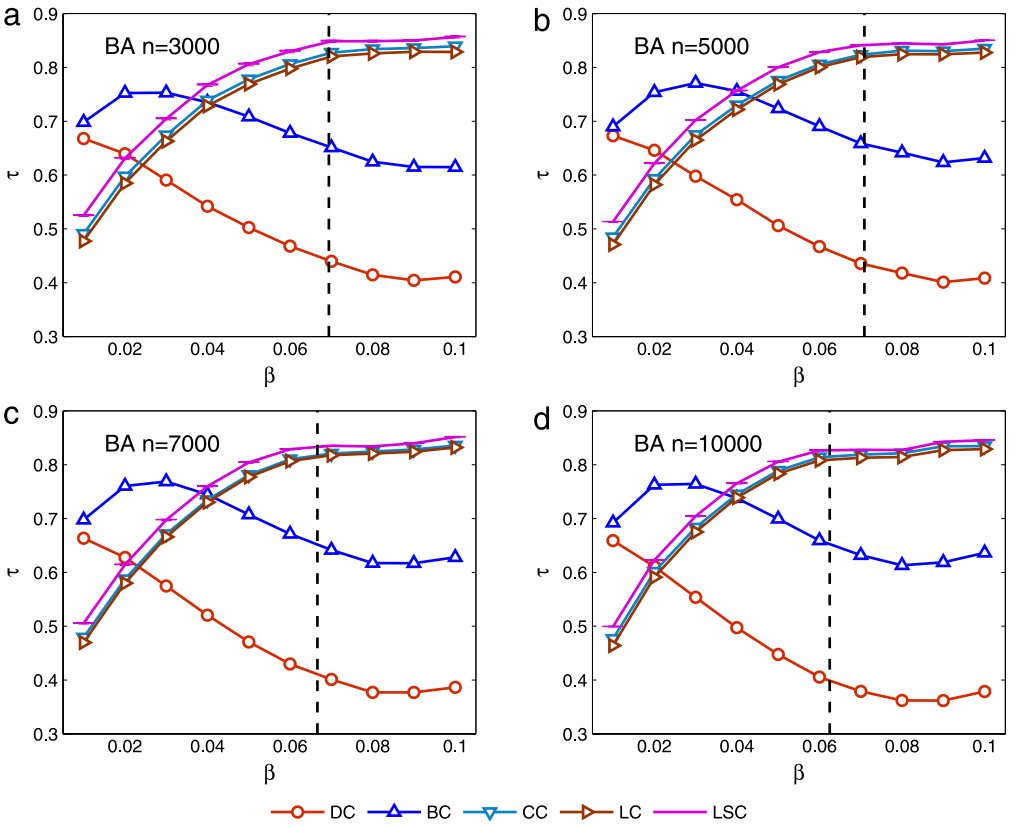

Fig. 5. (Color online) The Kendall's tau τ values for five centrality measures on four BA networks with $n = 3000, 5000, 7000, 10000$ and $m_t = 3$, where the vertical dashed line corresponds to the epidemic threshold β^c_{rand}. Centrality measures are degree centrality (DC), betweenness centrality (BC), closeness centrality (CC), local centrality (LC) and our local structural centrality (LSC). We show the performance of our LSC method with α from 0 to 1 using box plots. Boxes cover 50% of data and whisker cover 95%. The line in a box represents the median of data. The performance of different centrality measures are similar across networks with different sizes and our LSC method achieves the best performance on a wide range of β. Also, since the clustering coefficients for these BA networks are too small (0.003–0.009), even α varies, the final centrality value of the target node does not change too much, hence the performance of our method slightly changes with the varying of α. The results are obtained by averaging over 1000 independent realizations where the spreading probability β is ranging from 0.01 to 0.1.

Table 5
The distinct D values for DC, KS, LC and LSC ($\alpha = 0.2$) on all four real networks.

	Email (%)	Blog (%)	PGP (%)	Twitter (%)
DC	4.24	1.36	0.78	0.16
KS	0.97	0.18	0.24	0.05
LC	96.29	46.08	41.52	26.13
LSC	**97.62**	**57.36**	**63.46**	**81.05**

It is worth noting that, though the k-shell method is claimed to be able to identify the most influential nodes, it can only achieve small $\langle\hat{\tau}\rangle$ values when the number of the most influential nodes L is small in Fig. 8. We argue the main reason for these small $\langle\hat{\tau}\rangle$ values is that the k-shell method assigns same k-shell centrality value to nodes on the same shell and then fails to distinguish the spreading ability of the most influential nodes. Nodes pair with the same centrality value is neither concordant nor discordant and it contributes zero to calculate n_c and n_d in Eq. (8). To further explain our argument, we show the top-10 influential nodes by averaging spreading ability and their ranks by all centrality measures in Table 6. Here, nodes with the same centrality value are assigned with the same rank. We can clearly see that, for k-shell centrality measure, the top-10 nodes share the same rank in Email and Blog networks. In PGP and Twitter networks, the k-shell centrality measure performs better and has 3 and 7 ranks for the top-10 nodes respectively. All of the other centrality measures can well distinguish the spreading ability of top-10 nodes. Also, we can observe that our proposed local structural centrality and local centrality can perform competitively good and give more accurate ranks for top-10 nodes than other measures considered. Betweenness centrality measure performs the worst since the most influential nodes cannot get higher ranks by this measure. We will further examine the ability of a certain centrality measure to distinguish the spreading ability of nodes in Section 5.5.

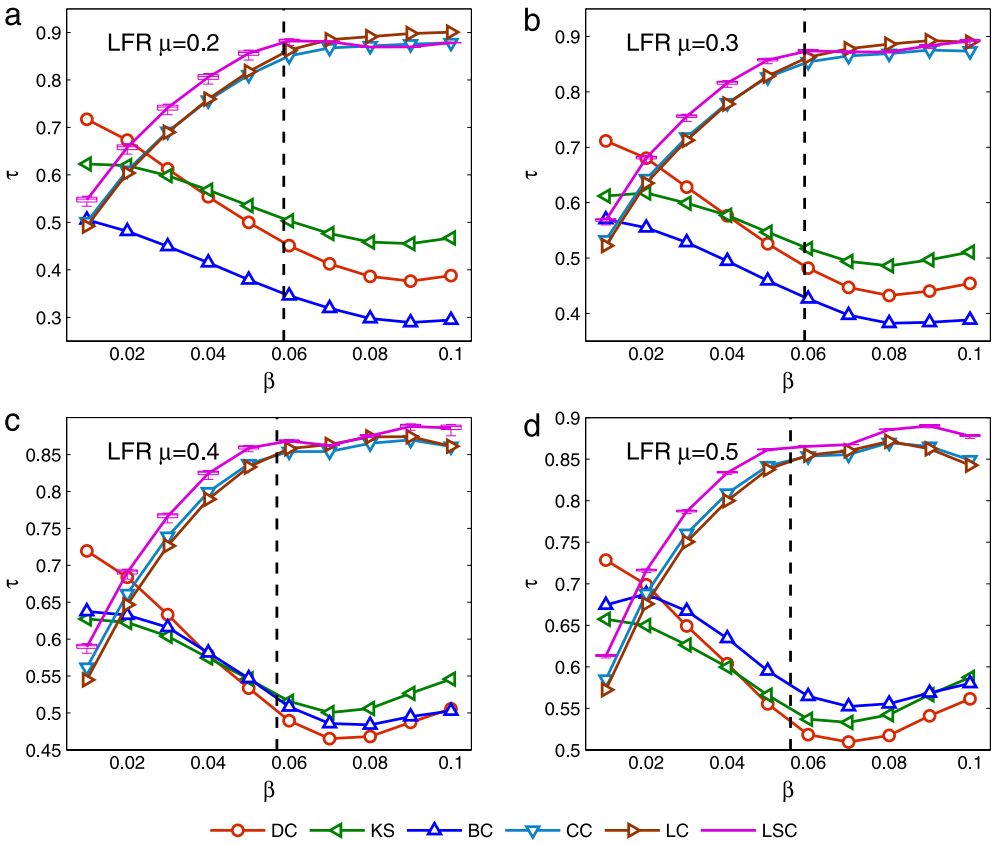

Fig. 6. (Color online) The Kendall's tau τ values for all six centrality measures on four LFR networks with $n = 5000$ and $\mu = 0.2, 0.3, 0.4, 0.5$, where the vertical dashed line corresponds to the epidemic threshold β^c_{rand}. Centrality measures are degree centrality (DC), k-shell centrality (KS), betweenness centrality (BC), closeness centrality (CC), local centrality (LC) and our local structural centrality (LSC). We show the performance of our LSC method with α from 0 to 1 using box plots. Boxes cover 50% of data and whisker cover 95%. The line in a box represents the median of data. From which we can see that our LSC method can always achieve the best performance when the spreading probability β is around the epidemic threshold β^c_{rand}. Note that the size of box at each data point is larger than that of BA networks due to the fact that the clustering coefficients for these LFR networks are relatively larger (0.09–0.42) and the balance parameter α begins to take effect. The results are obtained by averaging over 1000 independent realizations where the spreading probability β is ranging from 0.01 to 0.1.

Since the Kendall's tau coefficient can only reveal whether a certain centrality measure can estimate the right order of the spreading ability of nodes, it cannot evaluate the real spreading ability of top-ranked nodes. Thus, we introduce another measurement, $\langle \bar{s} \rangle$, which is the mean of average spreading ability of the top-L nodes as ranked by each centrality measure, to examine the real spreading ability of top-ranked nodes. We compare the $\langle \bar{s} \rangle$ values of our proposed method with those of other centrality measures considered in four real networks. To illustrate the results more clearly, we show the comparison with local metrics including DC and LC in Fig. 9 and the comparison with global metrics including BC, CC and KS is presented in Fig. 10. The curve for a good centrality measure should be downward sloping, namely the mean of average spreading ability of top-L nodes decreases with the increasing of L. Obviously, we can see in Fig. 9 that our proposed centrality measure can achieve comparable performance with local centrality on all four networks and our method performs slightly better than local centrality when L is small. Both of them outperform degree centrality on entire range of L. Fig. 10 shows that our method performs the best in Email, Blog and PGP networks, while k-shell centrality measure achieves competitively good performance in Twitter when L grows larger. That indicates k-shell centrality measure can well identify a set of influential nodes, though it cannot rank the spreading ability of these influential nodes accurately. Generally speaking, by comparing $\langle \hat{\tau} \rangle$ and $\langle \bar{s} \rangle$, we can conclude that our local structural centrality can better identify the influential nodes and can better rank the spreading ability of the influential nodes.

5.3. Performance on artificial networks

Besides the real networks, we also investigate the performance of the proposed LSC method on artificial networks generated by two kinds of network models, the classic Barabási–Albert (BA) network model [26,5] and the Lancichinetti–Fortunato–Radicchi (LFR) network model [27].

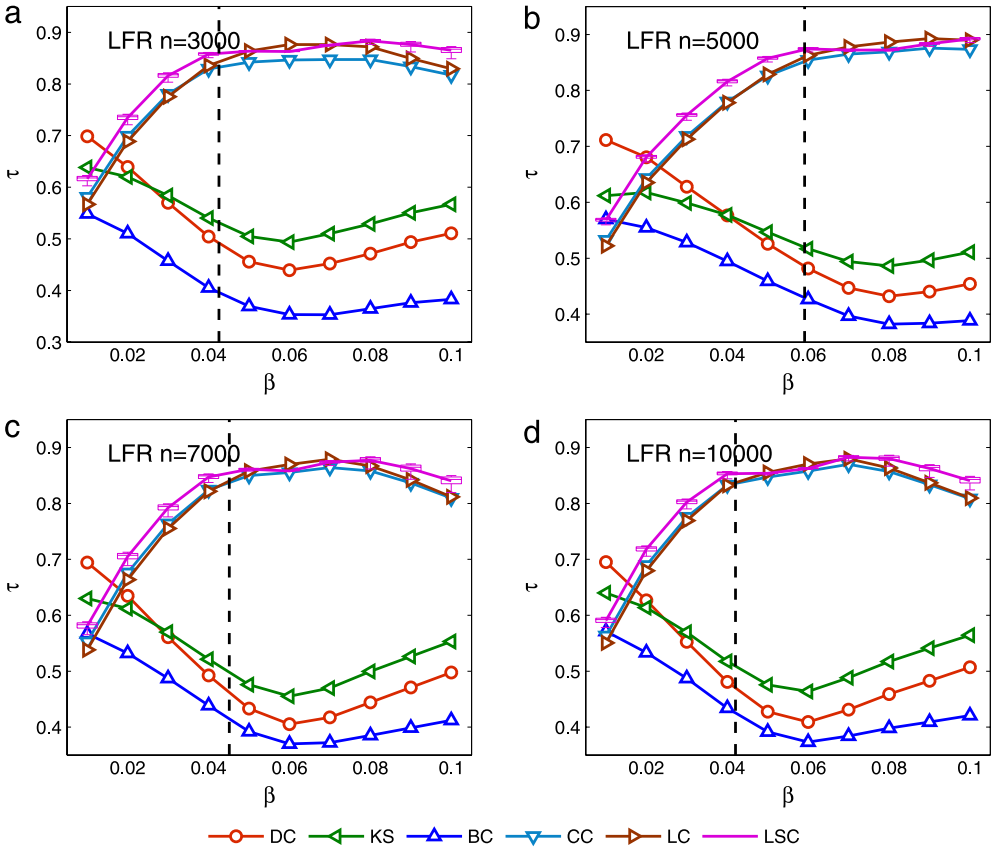

Fig. 7. (Color online) The Kendall's tau τ values for all six centrality measures on four LFR networks with $n = 3000, 5000, 7000, 10000$ and $\mu = 0.3$, where the vertical dashed line corresponds to the epidemic threshold β^c_{rand}. Centrality measures are degree centrality (DC), k-shell centrality (KS), betweenness centrality (BC), closeness centrality (CC), local centrality (LC) and our local structural centrality (LSC). We show the performance of our LSC method with α from 0 to 1 using box plots. Boxes cover 50% of data and whisker cover 95%. The line in a box represents the median of data. The performance of different centrality measures are similar across networks with different sizes and our LSC method achieves the best performance on a wide range of β. Also, the size of box at each data point is larger than that of BA networks due to the fact that the clustering coefficients for these LFR networks are relatively larger (around 0.27) and the balance parameter α begins to take effect. The results are obtained by averaging over 1000 independent realizations where the spreading probability β is ranging from 0.01 to 0.1.

The Barabási–Albert (BA) network model generates a random scale-free network with the preferential attachment mechanism: Initially, there are m_0 nodes in the network. Then, at each time step, a node with m_t links will connect to the existed nodes with a probability that is proportional to the number of links that the existing nodes already have. Fixing the number of nodes $n = 5000$ and $m_0 = 20$, we generate four BA networks by varying the number of attachments $m_t = 3, 4, 5, 6$. Note that, by decomposing the BA network using k-shell method, one can find that all nodes except the initial m_0 nodes have the same k-shell value, i.e. $C_{KS} = m_t$. That is to say the k-shell method cannot rank the spreading ability of nodes in BA networks. So we compare the performance of our proposed LSC method with the other four centrality measures and show the Kendall's tau values τ in Fig. 4. We use box plots to show the performance of our proposed method with α varying from 0 to 1 at each β. In box plots, boxes cover 50% of the data and whiskers cover 95%. The line in a box represents the median of the data. It can be seen from Fig. 4 that our LSC method achieves the best performance on a wide range of β on all four networks and the performance of our method is quite stable across all four networks. Since different m_t result in different degree distribution, the results in Fig. 4 indicate that the spreading process is significantly affected by the heterogeneity of degree distribution. Further, by fixing the number of attachments $m_t = 3$ and $m_0 = 20$, we generate four BA networks with sizes $n = 3000, 5000, 7000, 10000$. We show the performance of all centrality measures except k-shell method on four BA networks with different sizes in Fig. 5. We can see that the τ values for CC, LC and our LSC method increase with the increasing of spreading probability β and our LSC method shows the best performance when β grows larger. Also, by checking the performance of our LSC method across BA networks with different sizes, we find that the impact of the size of BA network on the performance of our LSC method is insignificant.

By examining each box in every subfigure of Figs. 4 and 5, we can find that the box is quite small at each data point, which indicates that the performance of the proposed method does not change too much with the varying of α. It is not surprising since the clustering property of the classic BA network with sufficient large size is not apparent [42]. The clustering coefficients of our generated BA networks are quite small (around 0.006), while the clustering coefficients for real networks

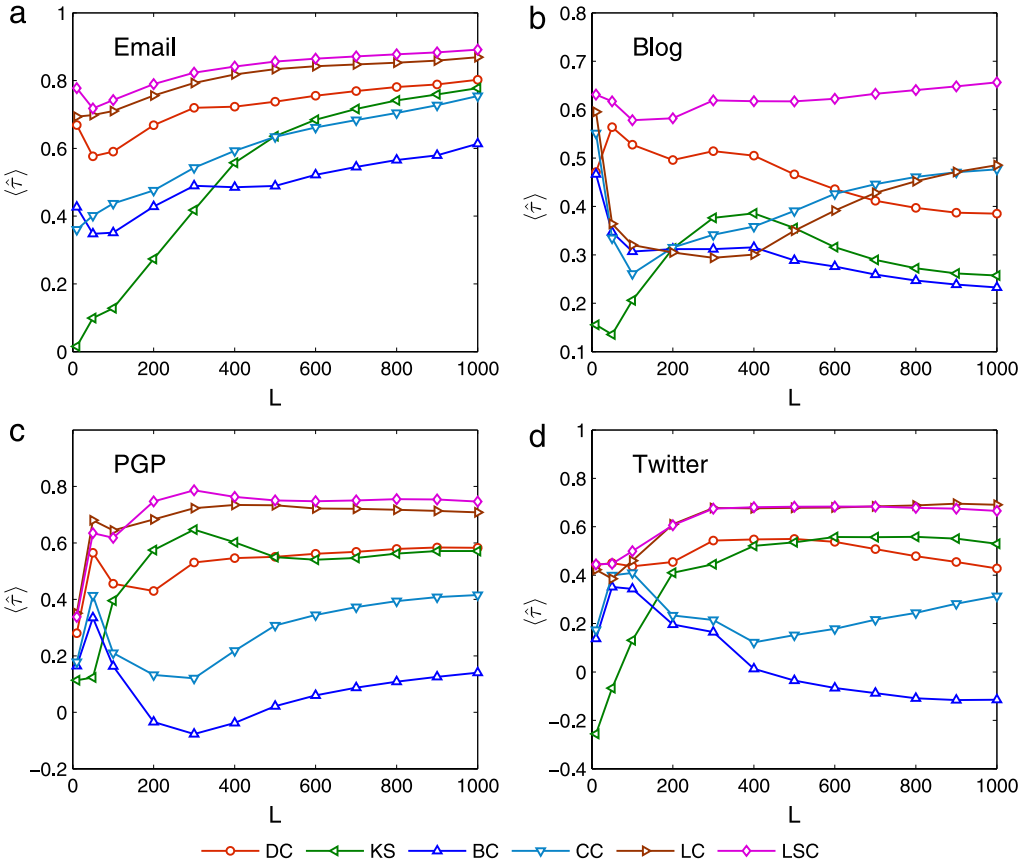

Fig. 8. (Color online) The $\langle\hat{\tau}\rangle$ values for all six centrality measures on four real networks when L varies from 10 to 1000. Centrality measures are degree centrality (DC), k-shell centrality (KS), betweenness centrality (BC), closeness centrality (CC), local centrality (LC) and our local structural centrality (LSC with $\alpha = 0.2$). L denotes the number of nodes with strongest average spreading ability considered when calculating $\langle\hat{\tau}\rangle$. The results are obtained by averaging over 1000 independent realizations.

are around 0.22 as shown in Table 3. Recall that in the explanation of our method in Section 3, giving a target node v, we split the contribution of its neighbor nodes into two parts in Eq. (7) and define the node's local clustering coefficient as its indirect contribution to the final centrality value of node v. When the clustering coefficient of the network is too small, most nodes in the network have small local clustering coefficients. Hence, even α varies, the final centrality value of the target node does not change too much. That explains why the performance of our method slightly changes with the varying of α. To sum up, our proposed LSC method shows better performance than other centrality measures considered on general BA networks and since the clustering coefficient of the BA network is too small, the performance our LSC method slightly changes with the varying of the balance parameter α.

The community structure is one of the most important features of complex networks [43,44]. Since our local structural centrality measure considers the local structure of a single node, one may wonder whether the community structure of the network could affect the performance of our method. Thus, we check the performance of our LSC method on networks with different community structure. To generate scale-free networks with different community structure, we use the Lancichinetti–Fortunato–Radicchi (LFR) network model [27]. In the LFR network model, both the degree and community size of the artificial network are assumed to follow power laws, with exponents γ and λ, respectively. The mixing parameter μ in LFR network model controls the community structure (modularity) of the generated network. The modularity decreases with the increasing of μ. Here, we fix the size of network to be 5000 which is large enough to consider the modularity of the complex networks. Then we vary the mixing parameter $\mu = 0.2, 0.3, 0.4, 0.5$ to generate four LFR networks with decreasing modularity. To eliminate the effects of other network properties, we fix the rest of the parameters in the LFR network model as follows: the average degree of nodes $\langle k\rangle = 6$, the max degree of nodes $k_{\max} = 70$, the power-law exponent of degree distribution $\gamma = 2$ and the power-law exponent of community size distribution $\lambda = 1$. The clustering coefficients for the four generated networks are 0.4223 ($\mu = 0.2$), 0.2692 ($\mu = 0.3$), 0.1009 ($\mu = 0.4$) and 0.0646 ($\mu = 0.5$). We show the Kendall's tau values τ for all the six centrality measures on four LFR networks in Fig. 6. Again we use the box plots to show the performance of our method with α varying from 0 to 1 at each β. We can see that when the spreading probability β is larger than 0.02, the performance of CC, LC and LSC are better than DC, KS and BC methods. Also, the performance of our LSC method is slightly better than CC and LC on a wide range of β. Since the interval between the epidemic threshold

Table 6
The top-10 nodes with the strongest spreading ability $\bar{s}$ and their corresponding ranks by degree centrality (DC), k-shell method (KS), betweenness centrality (BC), closeness centrality (CC), local centrality (LC) and our local structural centrality (LSC with $\alpha = 0.2$). $\bar{s}$ is obtained by averaging over 1000 implementations.

$\bar{s}$	DC	KS	BC	CC	LC	LSC
Email						
114.95	1	13	2	3	1	1
107.89	3	13	22	40	4	5
107.07	3	13	10	4	2	2
106.33	2	13	1	1	3	3
105.58	3	13	3	2	5	4
104.57	7	13	15	19	7	7
103.36	6	13	8	5	8	6
102.07	19	13	61	11	6	8
100.91	16	13	62	49	12	16
100.72	12	13	36	33	10	12
Blog						
22.83	1	12	3	3	1	1
16.50	15	12	45	11	2	4
15.47	16	12	50	16	5	3
13.44	3	12	16	7	3	2
13.02	37	12	66	15	4	7
12.73	4	12	18	10	11	6
12.52	2	12	22	31	46	5
12.31	47	12	173	43	21	8
11.67	57	12	214	45	24	12
11.42	62	12	253	54	20	14
PGP						
300.71	1	1	1	1	1	1
296.42	10	1	362	27	4	6
295.60	13	1	776	131	7	8
295.31	6	44	32	6	3	3
295.07	11	1	378	44	6	7
294.98	5	1	22	11	5	4
293.77	2	127	3	2	2	2
293.47	24	1	1 319	115	10	11
293.35	21	1	492	71	9	9
292.10	36	1	157	18	8	10
Twitter						
1139.25	1	48	14	11	1	1
1138.62	3	160	17	20	4	4
1134.64	5	160	30	18	6	5
1132.92	2	92	27	14	2	2
1132.81	7	1	7	2	3	6
1132.79	12	48	19	21	44	12
1131.37	23	1	35	26	10	11
1130.10	3	160	22	12	5	3
1129.22	17	1	52	25	7	9
1128.78	85	1	45 965	424	34	17

and a relatively small spreading probability is the critical area that is important to study, we can see that our LSC method always performs better in this area than other centrality measures considered. Also, it is worth noting that with the increasing the mixing parameter μ, the LC method becomes less effective on large β values. The reason may lie in the fact that for a larger μ, the network is lower clustered, considering neighbor nodes within too many hops from the initially infected node could bring too much noise information and harm the performance of the LC measure. While the DC measure which considers only one hop from the initially infected node performs much worse when β grows larger, our method which considers neighbor nodes within 3-hops can be viewed as a compromise between DC and LC measures to some extent. Further, by fixing the mixing parameter $\mu = 0.3$, we vary the sizes of LFR networks $n = 3000, 5000, 7000, 10000$, and show the Kendall's tau values τ for all measures in Fig. 7.Similar conclusion can be drawn from Fig. 7 that our method shows the best performance on a wide range of spreading probability especially when the spreading probability is around the epidemic threshold. Also, we can see that, the performance of our method is quite stable across networks with different sizes. The community structure of the network can be seen as a macro-view of the network since it considers all nodes and the overall structure of the network. In contrary, the local structure considered in our proposed centrality measure mainly focuses on node's neighborhood and can be seen as a micro-view of the network. The community structure and the local structure are

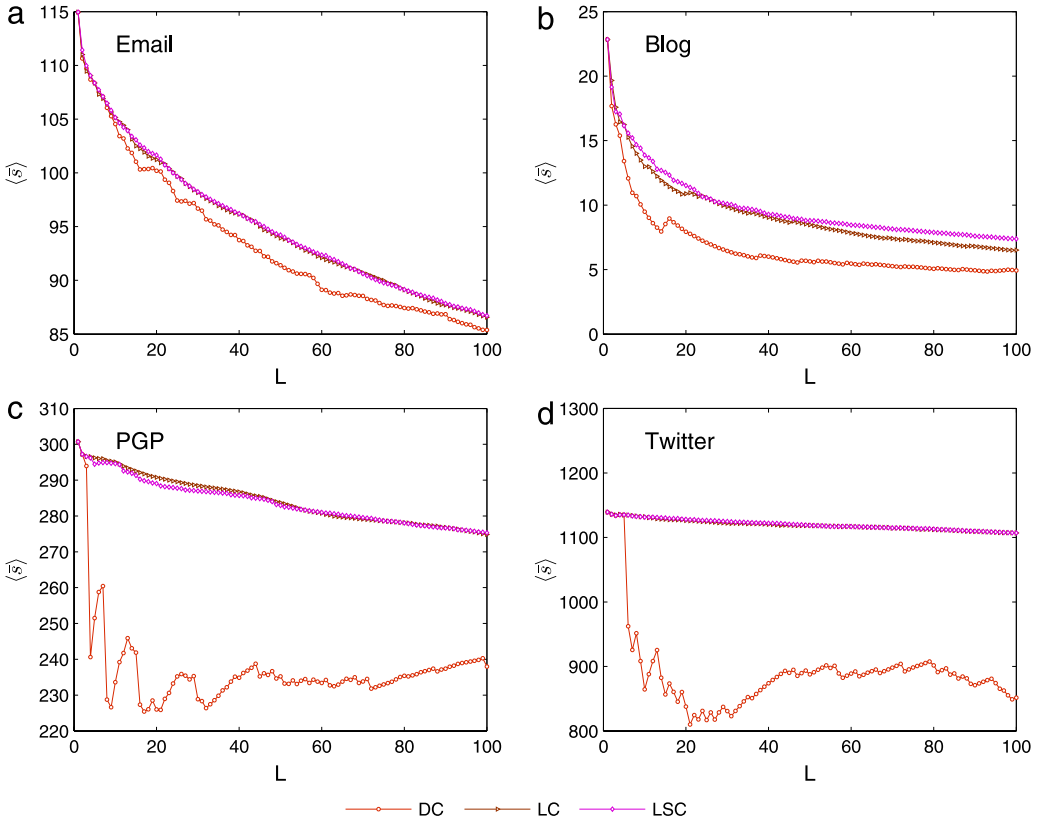

Fig. 9. (Color online) The mean of average spreading ability $\langle \bar{s} \rangle$ of top-L nodes as ranked by our local structural centrality (LSC with $\alpha = 0.2$) measures and local metrics including degree centrality (DC) and local centrality (LC). The results are obtained by averaging over 1000 independent realizations.

two kinds of views of the network. Through detailed simulations on LFR networks, we can see that our proposed method shows its effectiveness and robustness on networks with different community structure.

It is worth noting that the sizes of boxes in each subfigure of Figs. 6 and 7 are larger than these in Figs. 4 and 5. That indicates the performance of our LSC method varies with the balance parameter α on LFR networks. We argue that the clustering coefficients of the LFR networks with small μ are much larger than that of the BA networks, hence the centrality values in our LSC method change significantly with the varying of α. Since the optimal α for our LSC method changes with different spreading probability β, similar to the results on real networks, we observe that the LSC method can always achieve its best performance on a relatively small α value (smaller than 0.5).

To sum up, by checking the performance of our proposed method on artificial networks with different community structure and different sizes, we show that our method can always achieve the best performance on a wide range of spreading probability. That certifies the effectiveness and robustness of our proposed local structural centrality measure.

5.4. Performance on SI model

Besides the standard SIR model, we also use the standard SI model [25] to check the performance of our proposed LSC method. Compared with the SIR model, nodes in the SI model can only have two states: *susceptible* and *infected*. At each time step, the infected nodes infect their susceptible neighbors with probability β and remain infected. The spreading process ends when all the nodes in the network become infected. Here we denote the time that the spreading process ends as t_c. Note that no matter which node the epidemic spreading originates from, nodes in the network will all become infected in the end of spreading, i.e. $s_v(t_c) = n$, for every $v \in V$, where n is the number of nodes in the network. So we use the total number of infected nodes at time t ($t < t_c$), denoted by $s(t)$, to represent the spreading ability of the initially infected node. Larger $s(t)$ value indicates a stronger spreading ability. Here, we again set $\alpha = 0.2$ in our LSC method and compare the spreading ability of the nodes that either appear in the top-20 lists by local structural centrality or local centrality (not appearing in both lists) using SI model with $\beta = \beta^c_{rand}$ on four real networks. Note that, without considering the effect of common nodes in both ranking lists, the difference of these two methods can be well distinguished. We show the simulation results on four real networks in Fig. 11. To illustrate the difference between LC and LSC method more clearly, we limit the step of spreading t to 30. We can see from Fig. 11 that the curves which represent the local structural centrality are all whole above the curves

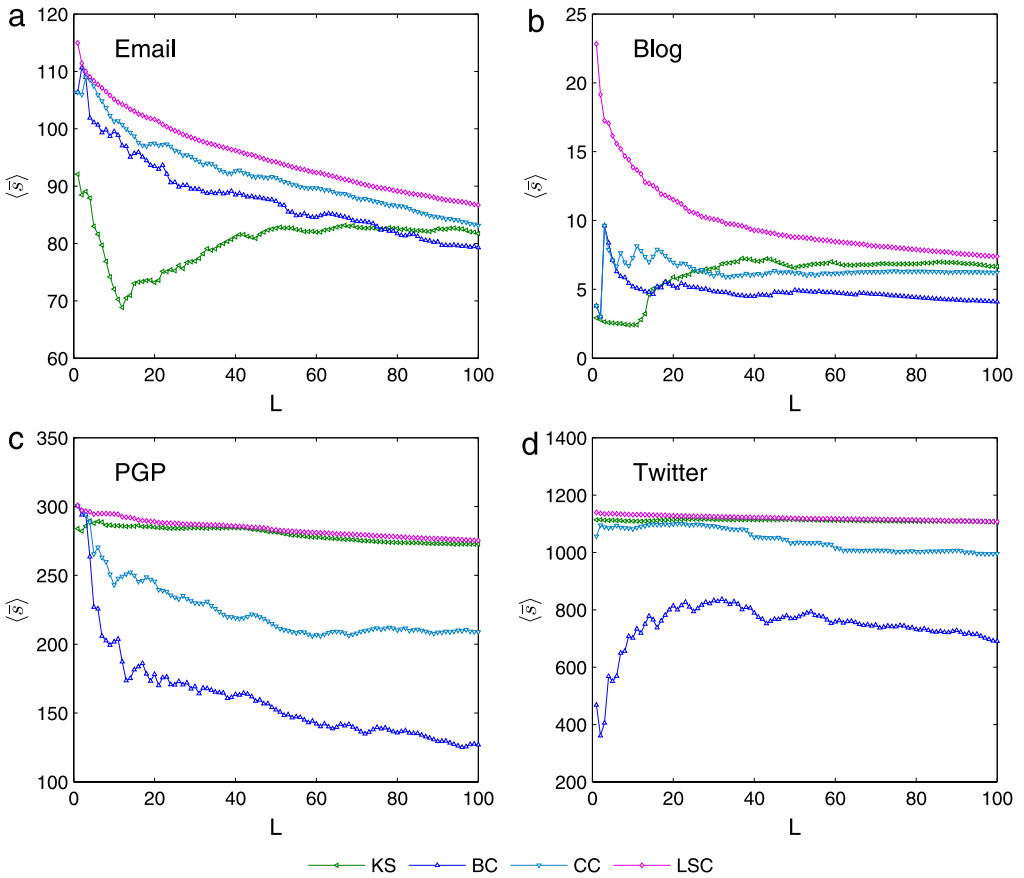

Fig. 10. (Color online) The mean of average spreading ability $\langle \tilde{s} \rangle$ of top-L nodes as ranked by our local structural centrality (LSC with $\alpha = 0.2$) measures and global metrics including k-shell centrality (KS), betweenness centrality (BC) and closeness centrality (CC). The results are obtained by averaging over 1000 independent realizations.

which represent the local centrality on all four networks. That indicates the nodes appearing in the top-ranked list of local structural centrality measure have stronger spreading ability and can spread the epidemic faster and wider, which further confirms the effectiveness of our proposed method.

5.5. Capability to distinguish nodes' spreading ability

Finally, we investigate the capability of centrality measures mentioned above to distinguish the spreading ability of the nodes. For a specific centrality measure, nodes in the network are ranked according to their centrality values in descending order. Nodes with the same centrality value have the same rank. The frequency of the appearance for a given rank is the number of nodes at this rank. We show the frequency of the appearance of different ranks in degree centrality, k-shell method, local centrality and our proposed local structure centrality on four networks in Fig. 12. Obviously, since both k-shell and degree centrality are coarse-grained ranking measures, they have only limited number of ranks and the number of each rank is quite large. Comparatively, local centrality and local structural centrality are fine-grained ranking measures and nodes are finer ranked in local structural centrality than in local centrality. Further, we define the distinct metric D as follows:

$$D = \frac{\text{number of distinct elements in } C}{n} \tag{9}$$

where C is the set of centrality values for all nodes in the network obtained by a certain centrality measure and n is the number of nodes in the network. The maximum value $D = 1$ indicates that all nodes in the network are assigned distinct centrality values and can be identically distinguished, while the minimum value $D = \frac{1}{n}$ indicates that all nodes are assigned the same centrality value. Obviously, a larger D indicates a finer rank of nodes. We show the results of D for all the real networks in Table 5. Our local structural centrality measure achieves the largest D values on all four networks. Hence, we can draw the same conclusion as drawn in Fig. 12 that our method can be much better at distinguishing the spreading ability of the nodes.

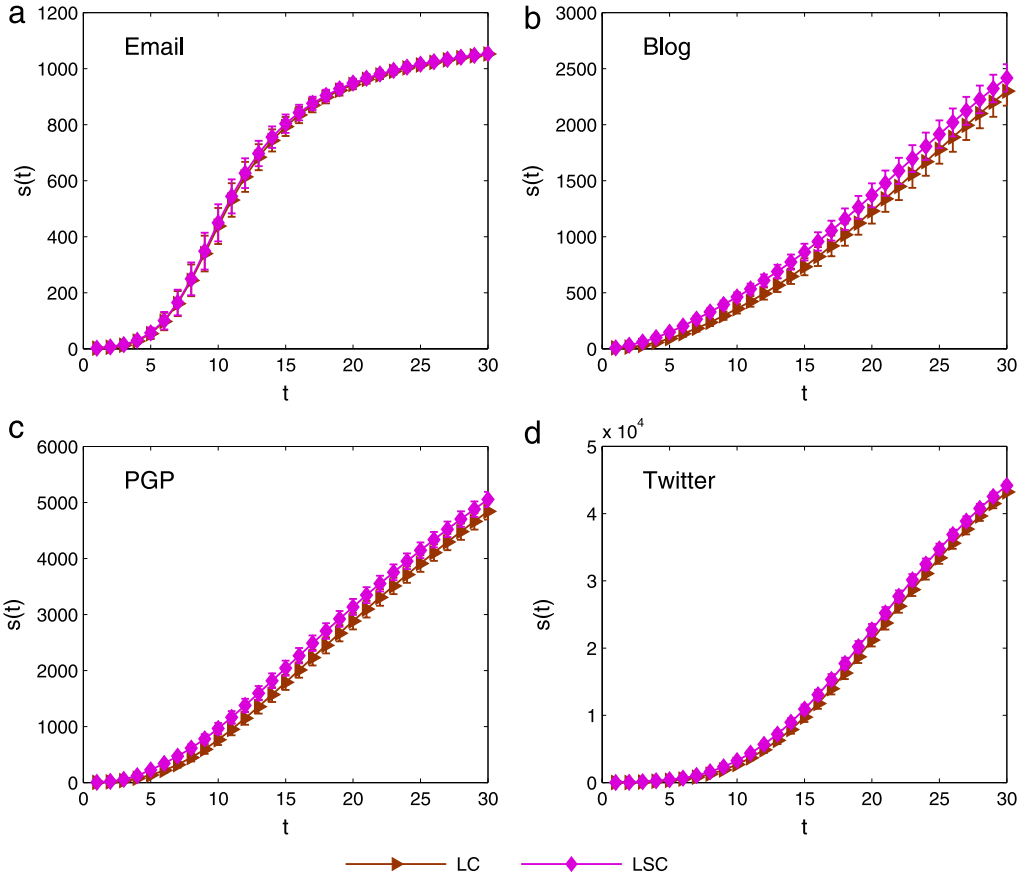

Fig. 11. (Color online) The number of infected nodes $s(t)$ as a function of time t under SI model, with the initially infected nodes being those either appearing in the top-20 list by local centrality (LC in brown right-pointing triangle) and local structural centrality (LSC with $\alpha = 0.2$ in magenta rhombus), but not appearing in both lists. The results are obtained by averaging over 1000 independent realizations where the spreading probability $\beta = \beta^c_{rand}$ on each network.

6. Conclusion and discussions

In this paper, we propose a local structural centrality measure to rank the spreading ability of nodes in the network. The proposed centrality measure considers not only the number of nearest neighbors of a node, but also the topological connections among the neighbors. To evaluate the performance, we apply our method on both artificial and real networks and use the SIR model to simulate the spreading process. By employing the Kendall' tau (τ) coefficient to measure the rank correlation between the ranked list generated by the simulation results and the ranked lists generated by different centrality measures, we find that our local structural centrality measure can rank the spreading ability of nodes more accurately than degree, k-shell, betweenness, closeness and local centrality. Further, by considering only the top-L nodes with strongest spreading ability, we show that our method can better rank the most influential nodes in the network. Through experiments on artificial networks generated by the Barabási–Albert (BA) network model and the Lancichinetti–Fortunato–Radicchi (LFR) network model, we show that our method can outperform other centrality measures in scale-free networks with different sizes and different community structure. We also use the SI model to simulate the spreading process and the simulation results further confirm the effectiveness of our proposed method. Finally, we investigate the capability of the centrality measures to distinguish the spreading ability of the nodes and show that our method performs better than other methods considered. The experiment results certify the effectiveness of our method. Also, since our method only considers node's local information, it is very efficient and can be applied into large scale networks.

Since the scale of social networks keep growing larger, the design of efficient and effective methods to rank the spreading ability of nodes in complex networks will be a long-term challenge. Due to the high computational complexity of global metrics, it is inconvenient to apply them into large scale networks. The k-shell method performs well in identifying a set of influential nodes, however it fails to distinguish the spreading ability of the most influential nodes. Thus, more and more attentions should be focused on designing local metrics and investigating the local structure around a node in the network. Our proposed method provides a simple insight on how the topological connections of the neighbors of a node affects its spreading ability. Further investigation could be done on analyzing the structure of nodes with multi-hops from the target

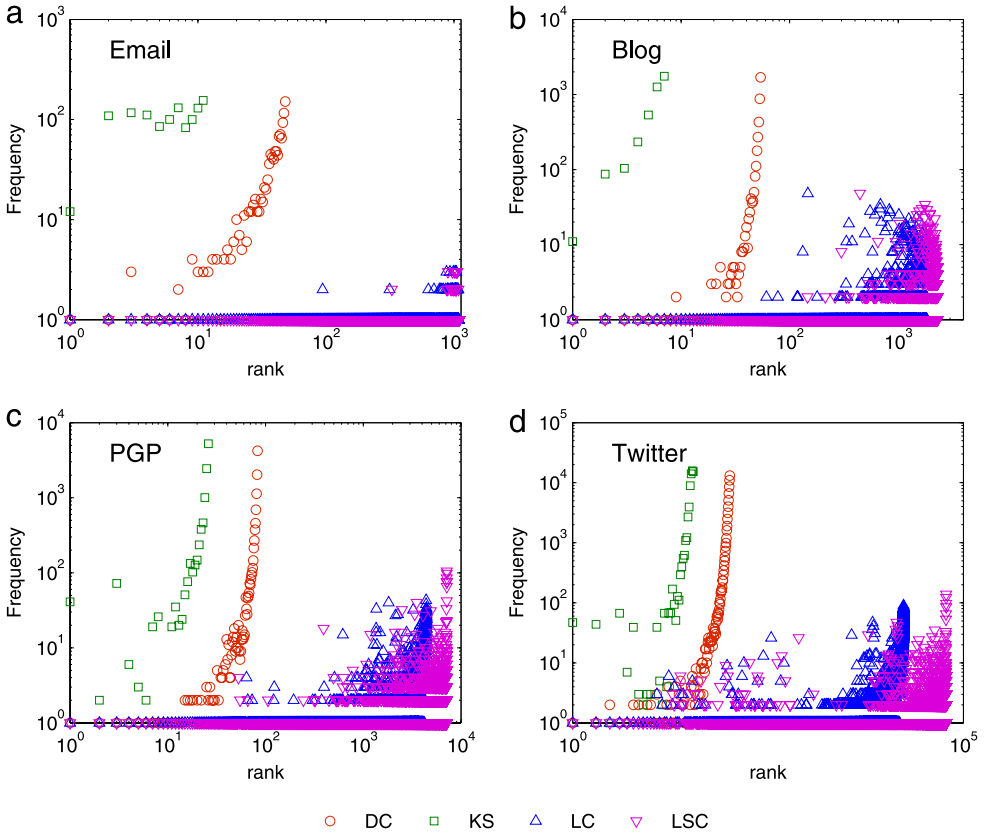

Fig. 12. (Color online) The frequency of appearance of different ranks in degree centrality (DC in red circle), k-shell centrality (KS in green square), local centrality (LC in blue upward-pointing triangle) and local structural centrality (LSC with $\alpha = 0.2$ in magenta downward-pointing triangle).

node. Also the community structure information [43] and the structural diversity [45] for the neighbors might be useful factors in estimating the spreading ability of a node.

By integrating more local structure information, we might devise more effective centrality measures that increase our understanding of the impact of node's local structure on its spreading ability. We believe that this paper may shed some light on the research of this direction.

Acknowledgments

We greatly appreciate the Editor's encouragement and the anonymous reviewer's valuable comments and suggestions to improve this work. This work is supported by the Natural Science Foundation of China (61272240, 61103151, 71301086, 11101243), the Doctoral Fund of Ministry of Education of China (20110131110028), the Natural Science foundation of Shandong province (ZR2012FM037) and the Excellent Middle-Aged and Youth Scientists of Shandong Province (BS2012DX017, BS2012SF016).

References

[1] R.M. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, Oxford and New York, 1991.
[2] J. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, Vol. 5, John Wiley & Sons, 2000.
[3] H. Hethcote, The mathematics of infectious diseases, SIAM Rev. 42 (4) (2000) 599–653.
[4] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 3200–3203.
[5] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.
[6] L. Lü, D.-B. Chen, T. Zhou, The small world yields the most effective information spreading, New J. Phys. 13 (12) (2011) 123005.
[7] M. Medo, Y.-C. Zhang, T. Zhou, Adaptive model for recommendation of news, Europhys. Lett. 88 (3) (2009) 38005.
[8] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, Rev. Modern Phys. 81 (2009) 591–646.
[9] J. Yang, C. Yao, W. Ma, G. Chen, A study of the spreading scheme for viral marketing based on a complex network model, Physica A 389 (4) (2010) 859–870.
[10] 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.
[11] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, T. Zhou, Identifying influential nodes in complex networks, Physica A 391 (4) (2012) 1777–1787.
[12] F. Bauer, J.T. Lizier, Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: a walk counting approach, Europhys. Lett. 99 (6) (2012) 68007.

[13] B. Hou, Y. Yao, D. Liao, Identifying all-around nodes for spreading dynamics in complex networks, Physica A 391 (15) (2012) 4012–4017.
[14] J.-G. Liu, Z.-M. Ren, Q. Guo, Ranking the spreading influence in complex networks, Physica A 392 (18) (2013) 4154–4159.
[15] D. Wei, X. Deng, X. Zhang, Y. Deng, S. Mahadevan, Identifying influential nodes in weighted networks based on evidence theory, Physica A 392 (10) (2013) 2564–2575.
[16] A. Zeng, C.-J. Zhang, Ranking spreaders by decomposing complex networks, Phys. Lett. A 377 (14) (2013) 1031–1035.
[17] L. Lü, Y.-C. Zhang, C.H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PLoS One 6 (6) (2011) e21202.
[18] Y.-B. Zhou, L. Lü, M. Li, Quantifying the influence of scientists and their publications: distinguishing between prestige and popularity, New J. Phys. 14 (3) (2012) 033033.
[19] L.C. Freeman, Centrality in social networks conceptual clarification, Social Networks 1 (3) (1978–1979) 215–239.
[20] G. Sabidussi, The centrality index of a graph, Psychometrika 31 (4) (1966) 581–603.
[21] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1) (1953) 39–43.
[22] A. Garas, F. Schweitzer, S. Havlin, A k-shell decomposition method for weighted networks, New J. Phys. 14 (8) (2012) 083030.
[23] L. Page, S. Brin, R. Motwani, T. Winograd, The pagerank citation ranking: bringing order to the web, in: Technical Report 1999-66, Stanford InfoLab, 1999, November.
[24] J.M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM 46 (5) (1999) 604–632.
[25] S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, Critical phenomena in complex networks, Rev. Modern Phys. 80 (2008) 1275–1335.
[26] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[27] A. Lancichinetti, S. Fortunato, F. Radicchi, Benchmark graphs for testing community detection algorithms, Phys. Rev. E 78 (2008) 046110.
[28] V. Batagelj, M. Zaversnik, An $o(m)$ algorithm for cores decomposition of networks, arXiv preprint cs/0310049.
[29] U. Brandes, A faster algorithm for betweenness centrality, J. Math. Sociol. 25 (2) (2001) 163–177.
[30] R.W. Floyd, Algortithm 97: shortest path, Commun. ACM 5 (6) (1962) 345.
[31] D.B. Johnson, Efficient algorithms for shortest paths in sparse networks, J. ACM 24 (1) (1977) 1–13.
[32] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (6684) (1998) 440–442.
[33] 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 (2003) 065103.
[34] S. Gregory, Finding overlapping communities using disjoint community detection algorithms, 2009, pp. 47–61.
[35] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, A. Arenas, Models of social networks based on social distance attachment, Phys. Rev. E 70 (2004) 056122.
[36] J. Hopcroft, T. Lou, J. Tang, Who will follow you back?: reciprocal relationship prediction, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM'11, ACM, New York, NY, USA, 2011, pp. 1137–1146.
[37] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (2002) 208701.
[38] H.-B. Hu, X.-F. Wang, Unified index to quantifying heterogeneity of complex networks, Physica A 387 (14) (2008) 3769–3780.
[39] C. Castellano, R. Pastor-Satorras, Thresholds for epidemic spreading in networks, Phys. Rev. Lett. 105 (2010) 218701.
[40] M.G. Kendall, A new measure of rank correlation, Biometrika 30 (1–2) (1938) 81–93.
[41] K. Klemm, M.Á. Serrano, V.M. Eguíluz, M. San Miguel, A measure of individual role in collective dynamics, Scientific reports 2.
[42] A. Fronczak, P. Fronczak, J.A. Hołyst, Mean-field theory for clustering coefficients in Barabási–Albert networks, Phys. Rev. E 68 (2003) 046126.
[43] S. Zhang, R.-S. Wang, X.-S. Zhang, Identification of overlapping community structure in complex networks using fuzzy-means clustering, Physica A 374 (1) (2007) 483–490.
[44] R.-S. Wang, R. Albert, Effects of community structure on the dynamics of random threshold networks, Phys. Rev. E 87 (2013) 012810.
[45] J. Ugander, L. Backstrom, C. Marlow, J. Kleinberg, Structural diversity in social contagion, Proc. Natl. Acad. Sci. 109 (16) (2012) 5962–5966.