Physica A 395 (2014) 549–559

Contents lists available at ScienceDirect

Physica A

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

Identifying and ranking influential spreaders in complex networks by neighborhood coreness

Joonhyun Bae *, Sangwook Kim

School of Computer Science and Engineering, Kyungpook National University, Sankyuk 1370, Daegu, Republic of Korea

HIGHLIGHTS

- The k-shell decomposition correctly identifies influential spreaders in complex networks.
- The monotonicity of the k-shell index is worse than other centrality measures.
- We propose an efficient ranking method by balancing the degree and the coreness of a spreader.
- The proposed method outperforms other measures in the scale-free network with community structure.

ARTICLE INFO

Article history:
Received 29 May 2013
Received in revised form 23 September 2013
Available online 4 November 2013

Keywords:
Complex network
Information diffusion
Centrality and ranking
k-shell decomposition

ABSTRACT

Identifying influential spreaders is an important issue in understanding the dynamics of information diffusion in complex networks. The *k-shell* index, which is the topological location of a node in a network, is a more efficient measure at capturing the spreading ability of a node than are the *degree* and *betweenness* centralities. However, the k-shell decomposition fails to yield the monotonic ranking of spreaders because it assigns too many nodes with the same k-shell index. In this paper, we propose a novel measure, *coreness centrality*, to estimate the spreading influence of a node in a network using the k-shell indices of its neighbors. Our experimental results on both real and artificial networks, compared with an epidemic spreading model, show that the proposed method can quantify the node influence more accurately and provide a more monotonic ranking list than other ranking methods.

1. Introduction

The problem of identifying influential spreaders in complex networks has gained increasing attention in the recent years [1–7]. The basic question is: how to measure the ability of a node to spread a message to a sufficiently large portion of the network. Centrality measures, such as the *degree* centrality, the *betweenness* centrality, and the *closeness* centrality, are studied in this literature [8,9].

Recently, Kitsak et al. [1] found that there are plausible circumstances under which the highly connected nodes or the highest-betweenness nodes have little effect on the range of a given spreading process. Instead, they showed that the most efficient spreaders are those located in the core of the network, which are identified using the *k-shell* (or *k-core*) decomposition method. However, the k-shell decomposition [10–16] tends to assign many nodes with an identical k-shell index, although the spreading capability of the nodes that reside in the same k-shell may differ from each other [5–7]. To overcome the shortcoming of monotonicity, Basaras et al. [5] proposed an alternative measure, *μ-power community index* (μ-PCI), which is an amalgam of the coreness and betweenness centralities. This metric balances the principles of

* Corresponding author. Tel.: +82 10 5105 3364; fax: +82 53 950 7699.
E-mail address: joonion@knu.ac.kr (J. Bae).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.10.047

the betweenness centrality and the transition density implied by the coreness measure. Zeng et al. [6] also noted that the k-shell method considers only the links between the residual nodes, whereas the links that connect to the exhausted nodes are entirely ignored. They proposed the *mixed degree decomposition* (MDD) method, where both the residual degree and the exhausted degree are considered. More recently, Liu et al. [7] presented an improved ranking method to generate a more distinguishable ranking list in terms of the node distance to the network core which is defined as the node set with the highest k-shell values.

In this paper, we propose a novel influence measure, the *coreness centrality*, to quantify the spreading capability of a node using the coreness of its neighbors. Our approach is based on the idea that a powerful spreader has more connections to the nodes that reside in the core of network. The k-shell indices of a node's neighbors could be good indicators of its spreading ability. To evaluate the effectiveness of the proposed measure, we apply the *susceptible-infected-recovered* (SIR) model for investigating an epidemic spreading process [17–20]. Our experimental results show that the coreness centrality provides a highly monotonic ranking list and overcomes the degeneracy of the k-shell decomposition. Moreover, we will show that our method is significantly more accurate than other centrality measures by evaluating the rank correlations with Kendall's tau.

The remainder of this paper is structured as follows. We briefly review previous studies and describe our centrality measures in the Section 2. In Section 3, we apply the SIR model to evaluate the effectiveness of the proposed method both in real and synthetic networks. The conclusions are presented in Section 4.

2. Methods

We focus on unweighted, undirected, and simple networks in this paper. Let $G = (V, E)$ be a graph with $n = |V|$ vertices and $m = |E|$ edges. The degree k of a node v is denoted by $k(v)$. A *k-core* (or *k-shell*) is a maximal connected subgraph of G in which all vertices have degree at least k.The *coreness ks* of node v, denoted by $ks(v)$, indicates that node v belongs to a k-shell but not to any $(k + 1)$-shell. The k-shell decomposition method [11–17] classifies all nodes of G into k-shells by removing nodes iteratively as follows. First, we remove all nodes with degree $k = 1$, and assign the coreness value $ks = 1$ to the removed nodes. Then the pruning is repeated until there are only nodes with degree $k>1$. Next, we repeat the pruning process in a similar way for the nodes of degree $k = 2$ and assign the coreness value $ks = 2$. This procedure is repeated until all nodes of G are removed and assigned to one of the k-shells. At the end of the decomposition, each node is associated with its own ks index, which indicates the topological location of a node.

Kitsak et al. found that the highly connected nodes may have notably different k-shell indices and can be located either in the core or in the periphery of the network [1]. As an illustrative example, shown in Fig. 1, let us consider two nodes, a and p, with identical degree $k = 8$. Node a is located in the core of the network with the k-shell index $ks = 3$, whereas the node p is placed in the periphery of the network with the k-shell index $ks = 1$. Hence, the degree of a node is not necessarily related to the spreading capability, whereas the k-shell index predicts the spreading influence of that node more accurately. However, by definition, the k-shell decomposition classifies many nodes with different degrees into the same k-shell, and the resulting ranking list has too many ties. Let us consider four nodes, a, b, c, and d, in the core of the schematic network. The k-shell index cannot distinguish the influence of these core nodes, although their spreading power differs from one another.

In order to discriminate the spreading ability of nodes in the same k-shell, one can extend the k-shell method with the degree of nodes. The *extended k-shell method* ranks the nodes based on the k-shell value, then sort the nodes with the same k-shell value based on their degrees in descending order. Besides this simple extension, there were several researches to enhance the performance of influence ranking in complex networks.

The *mixed degree decomposition* method, which was proposed by Zeng et al. [6], alters the k-shell decomposition process by considering both the residual degree k_r and the exhausted degree k_e, as follows:

$$k_m(v) = k_r + \lambda \bullet k_e, \tag{1}$$

where λ is a tunable parameter between 0 and 1. Note that the MDD method coincides with the k-shell method when $\lambda = 0$, but it is equivalent to the degree centrality when $\lambda = 1$. By considering the removed links, the MDD method could distinguish the nodes in the same shell. However, the parameter λ should be determined by the structure of a given network, and it is difficult to find the optimal parameter λ to achieve better result. Moreover, the MDD method gives equal importance to the removed nodes regardless of whether they reside in the core or in the periphery of the network. We set the parameter $\lambda = 0.7$ in this paper, which is used in Ref. [6].

More recently, Liu et al. [7] improved the k-shell index in terms of the distance from a target node to the network core, which is defined as the set of nodes with the highest k-shell value. The spreading influences of the nodes with identical k-core values could be distinguished in the following method:

$$\theta(v) = (ks_{\max} - ks(v) + 1) \sum_{w \in S_c} d(v, w), \tag{2}$$

where $ks_{\max}$ is the largest k-shell value of a network, S_c is the set of core nodes with the highest k-shell index $ks_{\max}$, and $d(v, w)$ is the shortest distance from the node v to the node $w \in S_c$. Although this non-parametric method could discriminate

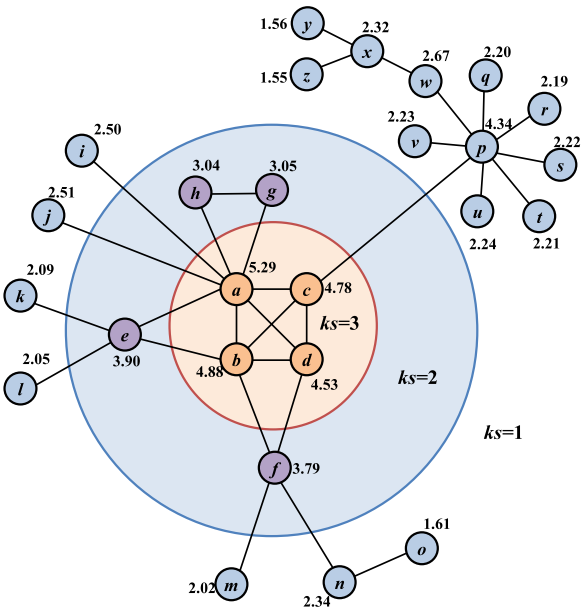


Fig. 1. A schematic network represented in Ref. [1]. There are two hubs a and p with identical degree $k = 8$, although a is located in the core of network $ks = 3$, and p is placed in the periphery of network $ks = 1$. Although k-shell decomposition efficiently detects the influential nodes, the nodes in the same shell are not distinguishable by the k-shell index. The spreading influence is represented around the nodes to show the different power of spreading ability of nodes.

the nodes in the same shell, the computational cost is relatively expensive by calculating the shortest distances to the core nodes. A more serious problem is that the shortest distance cannot be obtained if there are several non-trivial disconnected components. Moreover, the core nodes themselves could not be distinguished by this method.

Inspired by these previous studies, we encounter a ranking method that considers the degree and the coreness of a spreader simultaneously. The basic assumption is that a spreader with more connections to the neighbors located in the core of the network is more powerful. Based on this assumption, the *neighborhood coreness* C_{nc} of node v is defined as:

$$C_{nc}(v) = \sum_{w \in N(v)} ks(w), \tag{3}$$

where $N(v)$ is the set of the neighbors adjacent to node v and $ks(w)$ is the k-shell index of its neighbor node w. Recursively, the *extended neighborhood coreness* C_{nc+} of node v is defined as follows:

$$C_{nc+}(v) = \sum_{w \in N(v)} C_{nc}(w), \tag{4}$$

where $C_{nc}(w)$ is the neighborhood coreness of neighbor w of node v.

In Fig. 1, the nodes in the core are discriminated by the coreness $C_{nc}(a) = 17, C_{nc}(b) = 13, C_{nc}(c) = 10$, *and* $C_{nc}(d) = 11$. The spreading influences of nodes, which are measured by the SIR simulation, are represented around the nodes. One can observe that the order of these measures corresponds to that of the spreading power of nodes. Note that coreness centralities endow node p with different ranks. The rank of node p is fourth by the neighborhood coreness $C_{nc}(p) = 10$, whereas the rank induced by extended coreness is eighth with the value $C_{nc+}(p) = 17$. This observation implies that the extended coreness C_{nc+} gives more weights to the coreness than does the neighborhood coreness C_{nc}. The ranking orders revealed by the methods described here are summarized in Table 1. Note that the ranks of extended k-shell method are similar to that of the MDD method. This result implies that the MDD method is considered as a straight forward extension of k-shell method, though it has the advantage of tunable parameters.

We can argue that the coreness centrality is likely to be effective for detecting influential nodes, because it balances the degree and the coreness of a spreader. It is remarkable that the coreness centrality itself is a local measure, whereas the k-shell decomposition is a global method. The k-shell decomposition algorithm can be efficiently implemented with the linear time complexity of $O(m)$, where m is the number of edges in the network [9]. Hence we can argue that our method is likely to be more efficient than other time consuming measures such as the betweenness centrality and the closeness centrality in terms of computational cost.

3. Experimental results

In this section, we evaluate the effectiveness and the efficiency of the proposed coreness centrality. Twelve real networks are investigated with an epidemic spreading process, the SIR model. We compare the rank correlation using Kendall's tau,

Table 1
The order of spreaders is revealed by different ranking methods: degree centrality k, k-shell decomposition ks, extended k-shell method $ks+$, mixed degree decomposition k_m, improved method θ, betweenness centrality C_B, closeness centrality C_C, neighborhood coreness C_{nc}, and extended neighborhood coreness C_{nc+}.

Rank	k	ks	$ks+$	k_m	θ	C_B	C_C	C_{nc}	C_{nc+}
1	a, p	a, b, c, d	a	a, p	a, b, c, d	p	c	a	a
2	b	e, f, g, h	b	b	e, f	c	a, b, p	b	b
3	c, d, e, f	Others	c, d	c, d	g, h	a	d	d	c
4	x	–	e, f	e, f	i, j, p	f	e	c, p	d
5	g, h, n, w	–	g, h	x	k, l, m, n	w	f	e, f	e
6	Others	–	p	g, h, w	q–v, w	b	w	g, h	f
7	–	–	x	n	o	e, x	g, h	i, j, n, x	g, h
8	–	–	n, w	Others	x	d	i, j, q−v	k, l, m, w	p
9	–	–	Others	–	y, z	n	x	Others	i, j
10	–	–	–	–	–	Others	k, l, n	–	w
11	–	–	–	–	–	–	m	–	q–v
12	–	–	–	–	–	–	y, z	–	n
13	–	–	–	–	–	–	o	–	k, l, m
14	–	–	–	–	–	–	–	–	x
15	–	–	–	–	–	–	–	–	o, y, z

which considers the ties. We also apply the methods to artificial networks, the preferential attachment model and the LFR benchmark networks, to understand the impact of the degree distribution and the community structure of complex networks.

3.1. Evaluation methodologies

First, we define the *monotonicity* M of ranking vector R to quantify the resolution of different ranking methods, as follows:

$$M(R) = \left[1 - \frac{\sum_{r\in R} n_r(n_r-1)}{n(n-1)}\right]^2, \tag{5}$$

where n is the size of ranking vector R and n_r is the number of ties with the same rank r. This metric quantifies the fraction of ties in the ranking list. The monotonicity $M(R)$ is 1 if the ranking vector R is perfectly monotonic, and it becomes 0 if all nodes in R have an identical rank.

To evaluate the performance of the ranking methods, we use the *susceptible-infected-recovered* (SIR) model to examine the spreading influence of nodes [18–21]. In the SIR model, we set a node of interest v to be infected initially to investigate the influence of node v. All other nodes are set to be susceptible at the initial stage. Then, at each step, each infected node is recovered after it attempts to infect its susceptible neighbors with the *infection probability* β. This process is repeated until there remains no infected node in the network. The number of recovered nodes at the end of an epidemic process is an indicator to estimate the influence of the initially infected node v. The *spreading influence* $\sigma(v)$ of node v is defined as the number of recovered nodes averaged over a sufficiently large number of simulations. We set the number of simulations to be 1000 for small network and 100 for large networks $m > 10^4$.

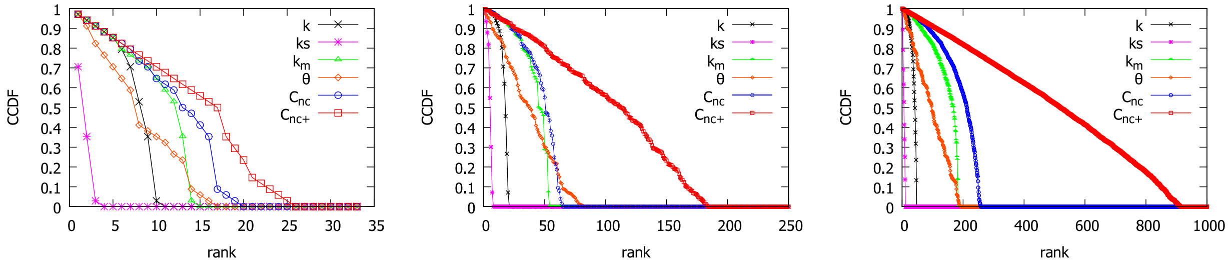
In order to quantify the correctness of the ranking methods, we adopt Kendall's tau as a rank correlation coefficient [21,22]. Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be a set of joint ranks from two ranking lists, X and Y, respectively. Any pair of ranks (x_i, y_i) and (x_j, y_j) is said to be *concordant* if the ranks for both elements agree with each other: that is, if $x_i > x_j$ and $y_i > y_j$ or if $x_i < x_j$ 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. Then, considering the ties, Kendall's tau τ of two rank vectors R_1 and R_2 is defined as follows:

$$\tau(R_1, R_2) = \frac{n_c - n_d}{\sqrt{(n_t - n_{t1})(n_t - n_{t2})}}, \tag{6}$$

where n_c and n_d are the numbers of concordant and discordant pairs respectively and $n_t = n(n-1)/2$, $n_{t1} = \sum_i t_i(t_i-1)/2$, $n_{t2} = \sum_j t_j(t_j-1)/2$, where n is the size of rank vectors and t_i and t_j are the number of tied values in the ith and jth groups of ties, respectively. This metric quantifies the similarity of the orderings of the measures when ranked by each quantity.

3.2. Applications to the real networks

Now we evaluate the performance of the ranking measures in twelve real networks with different sizes: Zachary's Karate Club [23], Lusseau's Bottlenose Dolphins [24], Jazz musicians network [25], C.elegans metabolic network [26], Coauthorship network of scientists (NetScience) [27], E-mail network of URV [28], Communication network of Blogs [29], Western States Power Grid [30], PGP network [31], Collaboration networks of astrophysics (CA-AstroPh) and condensed matter physics



(a) Karate Club. (b) NetScience. (c) E-mail.

Fig. 2. The distribution of ranks in three real networks: Karate Club, NetScience, and E-mail. The CCDF plot of k-shell index drops quickly, whereas the distribution of coreness centrality decreases monotonically. The improved method is worse than the degree centrality in the leftmost critical range of most central nodes.

(CA-CondMat) [32], and Enron e-mail network [33]. To be fair, only the largest connected component is considered in the network of NetScience, CA-AstroPh, CA-CondMat, and Enron.

The monotonicity M of different ranking methods is summarized in Table 2. The extended coreness C_{nc+} is the best measure, and the neighborhood coreness C_{nc} is also more competitive than the degree centrality and the k-shell method. Note that the monotonicity of extended k-shell method ($ks+$) is similar to that of the MDD method. To clarify the ranking distribution, we plot a complementary cumulative distribution function (CCDF), as shown in Fig. 2. This plot clearly shows the excellence of our method at distinguishing the spreading capability of the influential nodes. Let us pay attention to the leftmost part of the CCDF plot. The improved method decreases rapidly because the measure depends on the shortest distance from the core nodes. This problem is critical because the influence measure usually concentrates on the central nodes in the network. The MDD method performs well in terms of monotonicity. However, we will show that the ranking result of the MDD method degenerates to the degree centrality.

Next, we investigate the correctness of different measures by comparing the ranking result with the spreading influence σ, which is obtained from the SIR simulation. In the SIR simulation, we set the infection probability β to be sufficiently larger than the *epidemic threshold* $\beta_{th} \sim \langle k \rangle / \langle k^2 \rangle$ of the network, where $\langle k \rangle$ and $\langle k^2 \rangle$ denote the average degree and the second order average degree, respectively [7,20]. The values of the epidemic threshold β_{th} and the infection probability β are presented in the second and third columns of Table 3. Then, the rank correlation coefficient, which is Kendall's tau τ of Eq. (6), is summarized in Table 3. One can observe that our methods outperform the other methods in most cases by showing that the coreness centrality is highly correlated with the size of the infected population σ of the SIR process. Based on this observation, we plot the correlation of the influence measures in two networks, NetScience and E-mail, as shown in Fig. 3. In Fig. 3(a)–(h), we witness that the coreness centrality is significantly correlated with the infection rate, although it produces extremely monotonic relations with the spreading power of nodes in the network. In Fig. 3(i)–(l), we observe that the traditional measures, i.e., the betweenness centrality and the closeness centrality, have little relationship with the influence capability of the spreaders in an epidemic process.

In order to thoroughly understand the relation between the coreness centrality measure and the degree and the k-shell index of nodes, we plot a heat map of these rankings with the influence measure σ. The correlated heat ρ of two ranks is measured by:

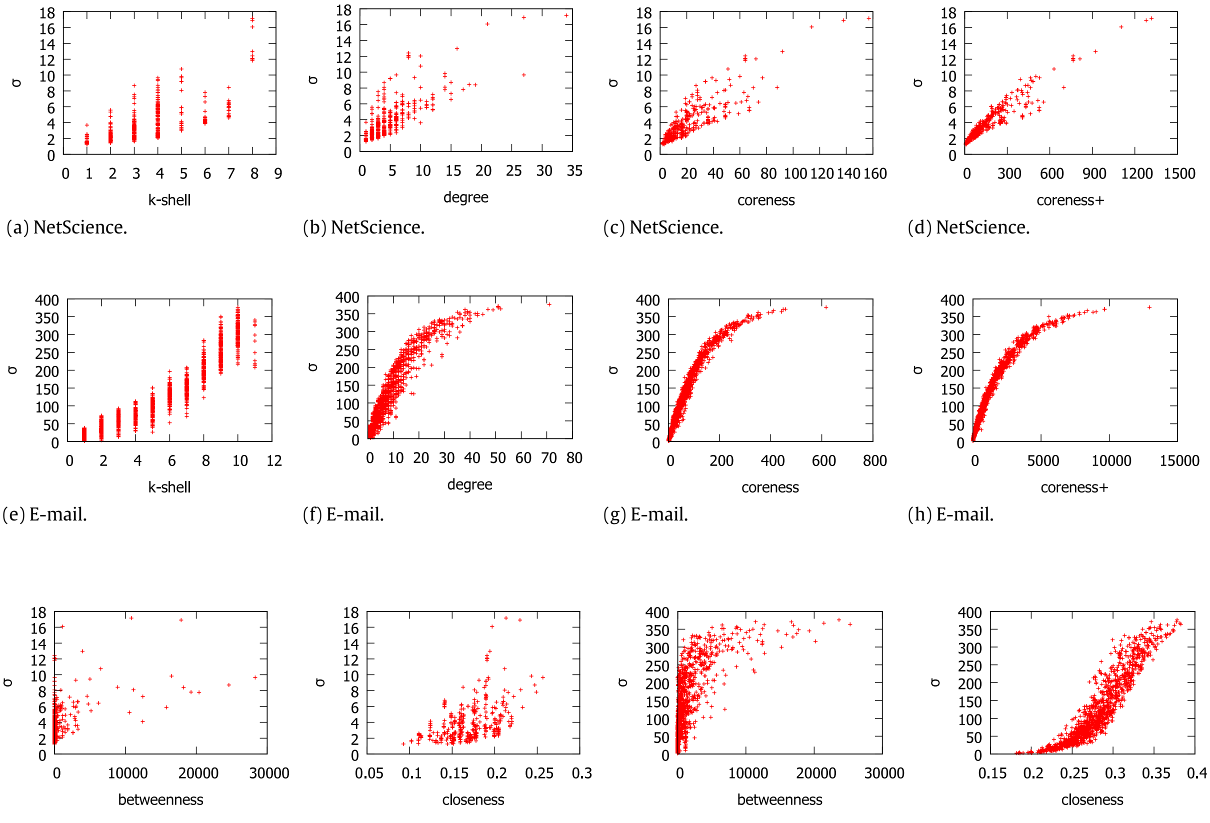
$$\rho(r_1, r_2) = \sum_{v \in S(r_1, r_2)} \frac{\sigma(v)}{|S(r_1, r_2)|}, \tag{7}$$

where $S(r_1, r_2)$ is the set of nodes with rank r_1 in one measure and rank r_2 in other measure. Fig. 4 shows the results when we compare the coreness centrality with the degree centrality and the k-shell index. In the NetScience network, there are a few hubs in the highest k-shell. However, in the E-mail network, there are powerful spreaders in the second highest shells with different degrees. From these results, we can argue that the coreness centrality is a better indicator to identify the spreading influence in complex networks.

To evaluate the effect of the infection probability β, we simulate the epidemic spreading on four networks by varying the value of β from 0.01 to 0.19. As shown in Fig. 5, the coreness centrality presents better result than the other measures, whenever the infection probability β is greater than the epidemic threshold β_{th}. Note that, in the E-mail network shown in Fig. 5(c), the rank correlation of the extended coreness rapidly decreases as the infection probability increases. Here, we could conjecture that the structural property of a network would affect the dynamics of information cascade over the networks. Hence we advance to study the impact of the characteristics of networks using artificial networks.

3.3. Applications to the artificial networks

The scale-freeness is one of the most frequently witnessed features of complex networks [34,35]. The Barabási–Albert (BA) model generates a random scale-free network with the preferential attachment mechanism: Initially, there are m_0



(a) NetScience. (b) NetScience. (c) NetScience. (d) NetScience.

(e) E-mail. (f) E-mail. (g) E-mail. (h) E-mail.

(i) NetScience. (j) NetScience. (k) E-mail. (l) E-mail.

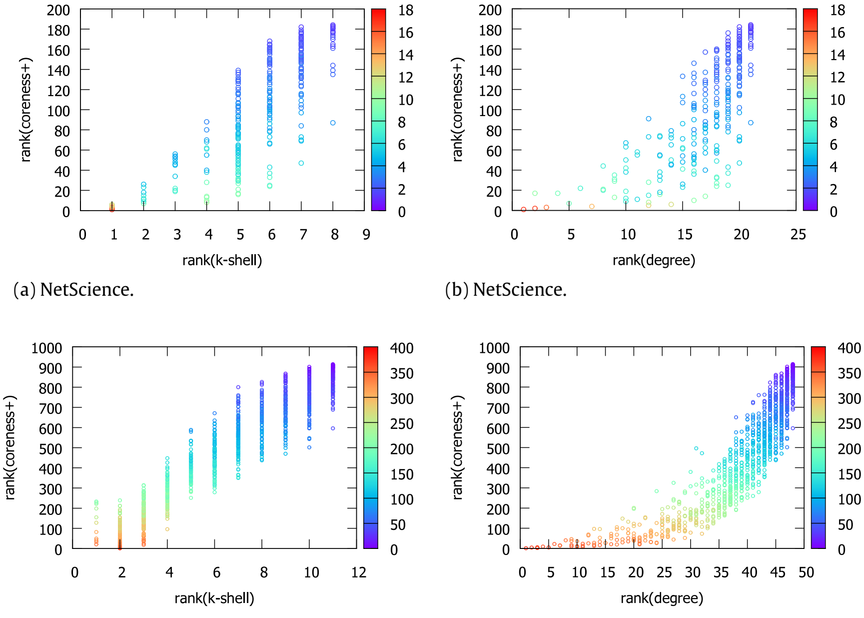
Fig. 3. The range of infection measured by the SIR simulation in the NetScience (a)–(d) and the E-mail (e)–(f) networks is plotted with the influence measures of different ranking methods. The coreness centrality is significantly correlated with the spreading influence of nodes in (c)–(f). The correlation of betweenness and closeness is worse than the degree centrality as shown in (i)–(l).

nodes in a network. Then a new node is added to the network and is connected to m_t existing nodes at each step t, with a probability that is proportional to the number of links of the existing nodes. The degree distribution that results from the BA model is scale-free, in particular, it is a power law of the form: $P(k) \sim k^{-3}$. Namely the *power-law exponent* γ of the degree distribution is $\gamma = 3$. We apply different ranking methods to the BA model by varying the number of attachments $m_t = 3, 4, 5$, using the same method as in Ref. [7]. We set the number of nodes $n = 1000$ and $m_0 = 10$. Fig. 6 depicts that the degree centrality, the MDD method [6], and the neighborhood coreness produces similar result, whereas the improved method [7] and the extended coreness present analogously better result. This result implies that the heterogeneity of the degree distribution significantly affects the spreading process of information diffusion. It is remarkable that the k-shell method does not yield a meaningful ranking in the BA model, as shown in Fig. 6(a), because most nodes are assigned into identical k-shell $ks = m_t$ in the BA model.

Here, we adopt a more complicated random network generation model, which is called the LFR benchmark [36]. In the LFR benchmark, the parameters enable to characterize the structural properties of complex networks. We set the parameters as follows: the number of nodes is $n = 1000$, the average degree of nodes is $\langle k \rangle = 10$, the maximum degree is $k_{max} = 50$, the power-law exponent of degree distribution is $\gamma = 2$, and the *mixing parameter of community structure* is $\mu = 0.2$. Then, we vary the average degree, the power-law exponent, and the mixing parameter to evaluate the effect of these prominent properties of complex networks.

First, we vary the average degree $\langle k \rangle = 5, 10, 15$, as shown in Fig. 7(a)–(c). When the infection probability is smaller than the epidemic threshold, the degree centrality is a good measure of the node importance in the information cascade. For a relatively large infection probability, the situation is identical. However, these cases are trivial in the analysis of epidemic spreading in complex networks. Instead, the interval between the epidemic threshold and a relatively small infection probability is the area that is important to study. One can observe that the coreness centrality performs better than the other measures in this critical section.

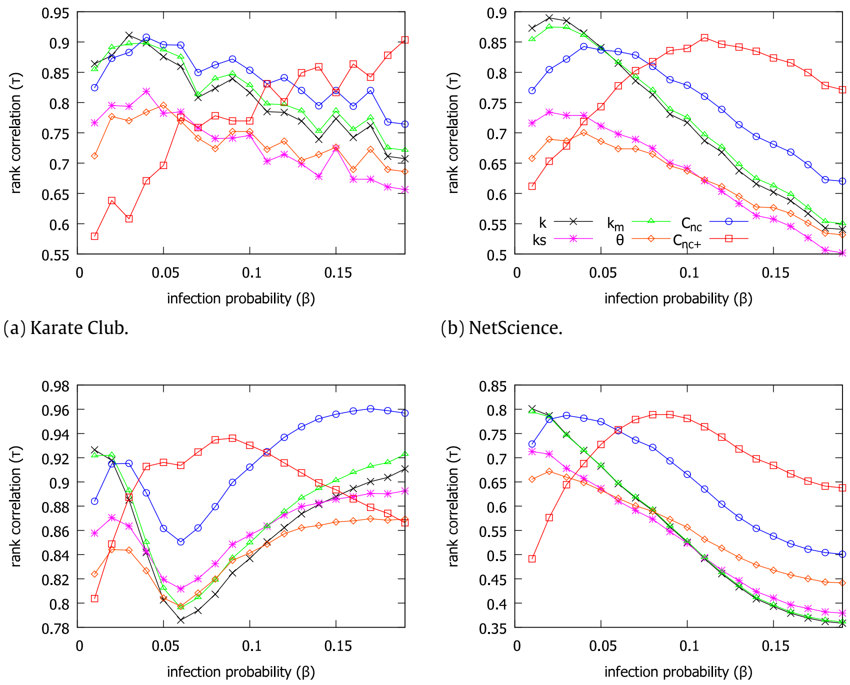
Next, we vary the power-law exponent of the degree distribution $\gamma = 2.0, 2.5, 3.0$, as shown in Fig. 8(a)–(c). Note that the neighborhood coreness approximates to the degree centrality as the power-law exponent increases. This result



(a) NetScience. (b) NetScience.

(c) E-mail. (d) E-mail.

Fig. 4. (Color online) The ranks of coreness centrality are compared with the k-shell index and the degree centrality in the NetScience and the E-mail networks. The correlated heat of each point is colored by the spreading influence ρ obtained by Eq. (7). In the NetScience network, (a) and (b), there are a few hubs in the highest k-shell, whereas there are influential spreaders in the second k-shell of the E-mail network. The color of each point clearly depicts the effectiveness of the coreness centrality measure.

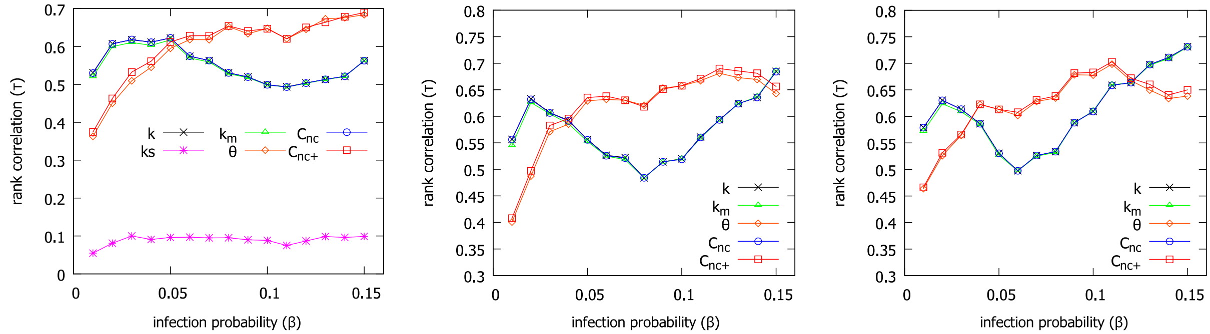


(a) Karate Club. (b) NetScience.

(c) E-mail. (d) Blogs.

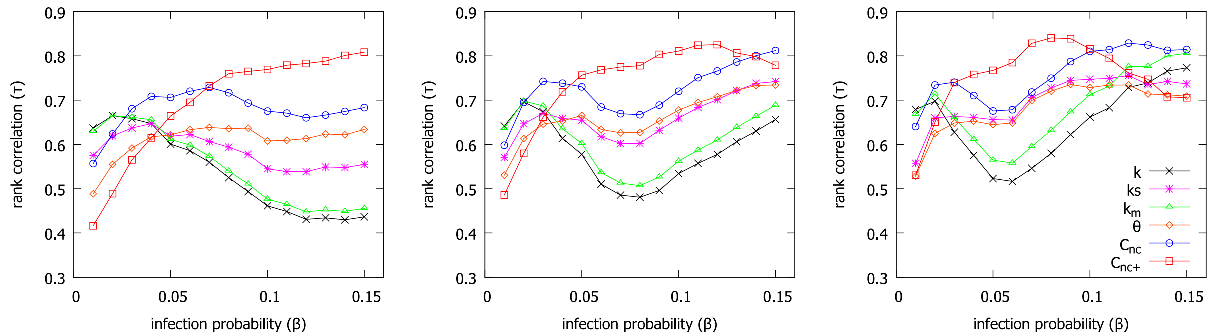
Fig. 5. The rank correlation coefficient, Kendall's tau τ, is plotted by varying the infection probability β in four networks: Karate Club, NetScience, E-mail, and Blogs. The coreness centrality outperforms other methods where the infection probability is larger than the epidemic threshold.

is plausible because the number of neighbors indicates the spreading ability of a node as the hubs have relatively large connections to their neighbors. In this case, only a small number of nodes reside in the core of the network and most nodes



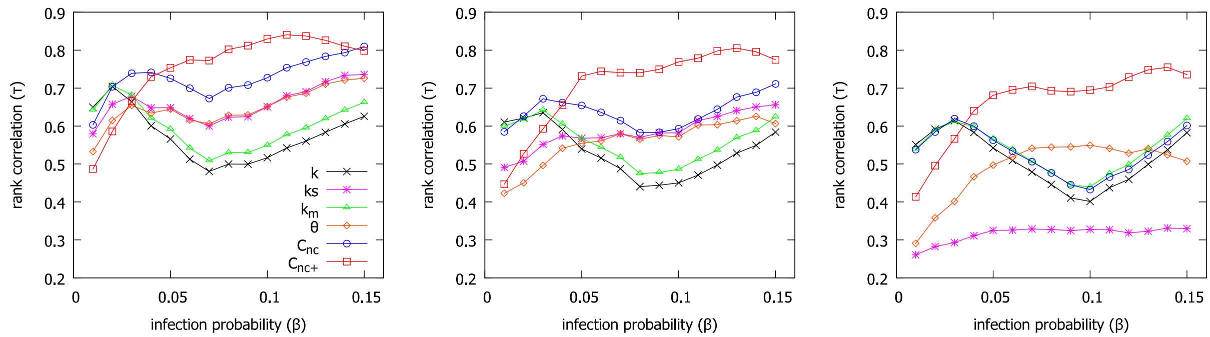
(a) BA ($mt = 3$). (b) BA ($mt = 4$). (c) BA ($mt = 5$).

Fig. 6. The variation of the rank correlation coefficient in the BA model: the degree centrality, the MDD method, and the neighborhood coreness show exactly identical performance in all the cases. In this random network, the extended coreness and the improved method show better performance. The k-shell method does not produce meaningful result because of the structural property of the BA model, where most nodes have an identical k-shell index.



(a) LFR ($\langle k \rangle = 5$). (b) LFR ($\langle k \rangle = 10$). (c) LFR ($\langle k \rangle = 15$).

Fig. 7. The variation of the rank correlation in the LFR benchmark network is plotted by changing the average degree $\langle k \rangle = 5, 10, 15$. The coreness centrality outperforms the other methods regardless of the densification of the networks. The variance of the MDD method is analogous with the degree centrality, whereas the other methods behave differently from the degree centrality.

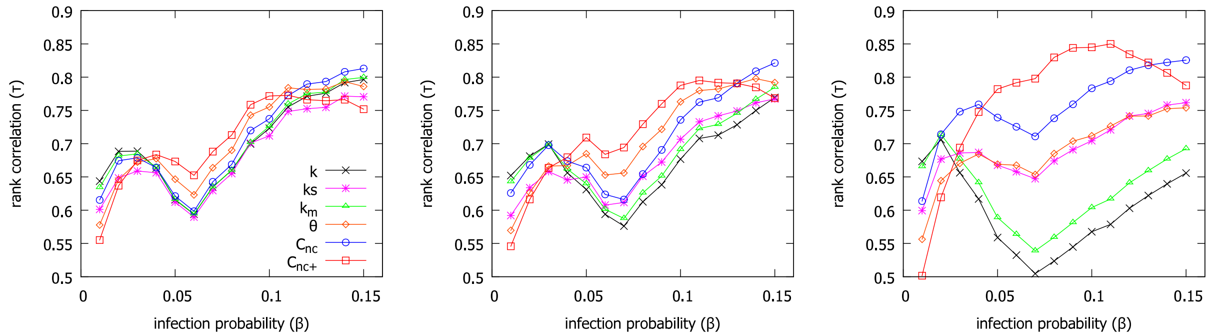


(a) LFR ($\gamma = 2.0$). (b) LFR ($\gamma = 2.5$). (c) LFR ($\gamma = 3.0$).

Fig. 8. The variation of the rank correlation in the LFR benchmark network is plotted by changing the power-law exponent $\gamma = 2.0, 2.5, 3.0$. When the heterogeneity of degree distribution grows, the performance of the k- shell method deteriorates significantly. Note that the neighborhood coreness approximates the degree centrality when the exponent increases. The extended coreness and the improved method show good performance regardless of the change of power-law exponent.

are located in the periphery of the network. This conjecture can be accommodated with the fact that the rank correlation of k-shell method rapidly deteriorates as the power-law exponent increases.

Next, we vary the mixing parameter $\mu = 0.8, 0.5, 0.2$, as shown in Fig. 9(a)–(c). The community structure is another famous property of complex networks that is rigorously studied in the last decade [37]. When the parameter is low, the LFR benchmark creates a random network with a significant community structure. Note that the difference among the ranking methods becomes obvious when the community structure is introduced into the network. This phenomenon is interpreted as follows: the clearer the community structure is, the better the coreness centrality identifies the influential spreaders in complex networks. It is reasonable in that the community leaders of relatively small clusters might be located



(a) LFR ($\mu = 0.8$). (b) LFR ($\mu = 0.5$). (c) LFR ($\mu = 0.2$).

Fig. 9. The variation of rank correlation in the LFR benchmark network is plotted by changing the mixing parameter $\mu = 0.8, 0.5, 0.2$. When the community structure becomes obvious, the difference among the rank correlations becomes clear. This result implies that the coreness centrality outperforms the other methods in complex networks with community structure. The improved method performs better than the neighborhood coreness in (a) and (b), where the community structure is unclear.

Table 2
The monotonicity M of different ranking methods was applied to twelve real networks. The values n and m are the number of nodes and links in each network, respectively. The value of M (•) is the monotonicity of the corresponding measures.

Network	n	m	$M(k)$	$M(ks)$	$M(ks+)$	$M(k_m)$	$M(\theta)$	$M(C_{nc})$	$M(C_{nc+})$
Karate Club	34	78	0.7079	0.4958	0.7413	0.7536	0.8791	0.8526	0.9472
Dolphins	62	159	0.8312	0.3769	0.8564	0.9041	0.9737	0.9284	0.9873
Jazz	198	2,742	0.9659	0.7944	0.9880	0.9882	0.9345	0.9982	0.9993
NetScience	379	914	0.7642	0.6421	0.8217	0.8215	0.9619	0.9302	0.9893
C.elegans	453	2,025	0.7922	0.6962	0.8693	0.8748	0.9582	0.9709	0.9975
E-mail	1,133	5,451	0.8874	0.8088	0.9204	0.9229	0.9783	0.9839	0.9991
Blogs	3,982	6,803	0.5654	0.4670	0.5885	0.5906	0.8794	0.8485	0.9868
PowerGrid	4,941	6,594	0.5927	0.2460	0.6600	0.6928	0.9604	0.7292	0.9419
PGP	10,680	24,316	0.6193	0.4806	0.6674	0.6678	0.9856	0.8919	0.9851
CA-AstroPh	17,903	196,972	0.9328	0.9142	0.9565	0.9558	0.9954	0.9955	0.9997
CA-CondMat	21,363	91,286	0.8616	0.8032	0.9085	0.9078	0.9914	0.9787	0.9986
Enron	33,696	180,811	0.7610	0.7389	0.7804	0.7803	0.9957	0.9688	0.9964

in the relatively lower k-shells than the community leaders of larger clusters. We can argue that the coreness centrality can capture the spreading capability of influential nodes when the community structure is obvious.

Finally, we examine the performance of ranking methods according to the degree of nodes. Since complex networks have power-law degree distribution, the most nodes have small degrees and a few hub nodes have large degrees that greatly exceed the average. In Table 4, we estimated Kendall's tau in the fraction of nodes divided by the order of degree by 10% using the LFR network with default parameters. As a result, the correlation between influence measure and other ranking methods is high in the group of large degree nodes and low in that of small degree nodes. Note that the coreness centrality shows significantly high performance in the group of small degree nodes. This result implies that our proposed methods can detect the difference of spreading ability with small links to the neighbor of nodes.

4. Conclusions

In this paper, we proposed an efficient ranking method, the *coreness centrality*, to measure the ability of spreaders in complex networks using the neighborhood coreness. The coreness centrality, which is estimated with the k-shell indices of the neighbors that are adjacent to a spreader, is a simple but notably powerful indicator to assess the capability of information dissemination through the network. Our approach is to consider both the degree and the coreness of a node coincidently by counting the k-shell indices of its neighbor or neighbor of neighbors. We evaluated the effectiveness of our method by comparing the ranking of coreness centrality with the size of the infected population in the SIR model. The experimental results suggest that the coreness centrality produces more monotonic ranking than the k-shell method and the degree centrality. Moreover, the ranking result of our method is highly correlated with the epidemic spreading range, compared with the other methods such as the betweenness centrality, the closeness centrality, the MDD method [6], and the improved method [7]. Using the experiments on the LFR benchmarks, we have found that the coreness centrality significantly outperforms the other measures in the scale-free networks with community structure.

Here, we focused on the unweighted and undirected networks. The generalization of the coreness centrality is worth studying because the ranking measure is independent to the k-shell decomposition method. Batagelj et al. [12] generalized the k-shell decomposition to directed networks. Garas et al. [14] also presented a weighted k-shell decomposition method. Our method can be applied with these generalized algorithms because of the locality of the measure. Once the k-shell index

Table 3
The correlation of different ranking measures, compared with the spreading influence σ obtained using the SIR simulation, is measured by Kendall's tau τ. β_{th} is an epidemic threshold of networks and β is the infection probability used in the SIR spreading process.

Network	β_{th}	β	$\tau(\sigma, k)$	$\tau(\sigma, ks)$	$\tau(\sigma, ks+)$	$\tau(\sigma, k_m)$	$\tau(\sigma, \theta)$	$\tau(\sigma, C_{nc})$	$\tau(\sigma, C_{nc+})$
Karate Club	0.129	0.15	0.7424	0.6777	0.7550	0.7557	0.6922	0.7940	0.8564
Dolphins	0.147	0.15	0.8061	0.7372	0.8011	0.8240	0.7332	0.8110	0.8777
Jazz	0.026	0.05	0.8733	0.8066	0.8481	0.8968	0.7659	0.9169	0.9178
NetScience	0.125	0.15	0.6039	0.5570	0.5733	0.6149	0.5489	0.6860	0.8339
C.elegans	0.025	0.05	0.6947	0.7509	0.7198	0.6997	0.8110	0.7930	0.8224
E-mail	0.054	0.10	0.8397	0.8580	0.8579	0.8524	0.8190	0.9148	0.9300
Blogs	0.072	0.10	0.5262	0.5235	0.5266	0.5281	0.5414	0.6690	0.7848
PowerGrid	0.258	0.30	0.5554	0.4909	0.5684	0.5770	0.3300	0.6188	0.7843
PGP	0.053	0.10	0.4806	0.4909	0.4850	0.4837	0.6024	0.6679	0.7219
CA-AstroPh	0.015	0.02	0.6985	0.7081	0.7023	0.7062	0.7223	0.7682	0.7986
CA-CondMat	0.045	0.06	0.5706	0.6035	0.5931	0.5819	0.6899	0.6771	0.7642
Enron	0.007	0.02	0.4692	0.4846	0.4790	0.4717	0.6584	0.6749	0.6671

Table 4
The performance of ranking methods according to the degree of nodes is tested in the LFR network with default parameters. Kendall's tau between the spreading influence σ and ranking methods is estimated in the fraction of nodes divided by the order of degree by 10%.

	Fraction of nodes ordered by degrees									
	0%–10%	10%–20%	20%–30%	30%–40%	40%–50%	50%–60%	60%–70%	70%–80%	80%–90%	90%–100%
$\tau(\sigma, k)$	0.517	0.230	0.044	0.072	0.149	0.141	−0.012	0.117	0.141	0.141
$\tau(\sigma, ks)$	0.761	0.781	0.750	0.602	0.597	0.518	0.409	0.203	0.141	0.141
$\tau(\sigma, k_m)$	0.604	0.622	0.585	0.438	0.477	0.534	0.373	0.164	0.141	0.141
$\tau(\sigma, \theta)$	0.713	0.734	0.643	0.571	0.642	0.465	0.438	0.398	0.458	0.430
$\tau(\sigma, C_{nc})$	0.814	0.812	0.787	0.749	0.754	0.810	0.791	0.788	0.803	0.827

is provided by decomposing the network globally, the coreness centrality can be obtained as a local measure. This feature is significant when we analyze notably large networks or highly dynamic networks. The computational complexity of k-shell decomposition is linear, i.e., it can be implemented with the time complexity of $O(m)$ [13]. Fortunately, researchers have recently developed a distributed algorithm [15] and a streaming algorithm [16], enabling the application of the coreness centrality to real-time applications [5]. Our future direction is to apply the coreness centrality to understand the dynamics of information diffusion in large-scale dynamic networks.

Acknowledgment

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2012M3C4A7032182).

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] 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.
[3] X. Zhang, J. Zhu, Q. Wang, H. Zhao, Identifying influential nodes in complex networks with community structure, Knowl.-Based Syst. 42 (2013) 74–84.
[4] B. Hou, Y. Yao, D. Liao, Identifying all-around nodes for spreading dynamics in complex networks, Physica A 391 (15) (2012) 4012–4017.
[5] P. Basaras, D. Katsaros, L. Tassiulas, Detecting influential spreaders in complex, dynamic networks, IEEE Comput. 46 (4) (2013) 24–29.
[6] A. Zeng, C.J. Zhang, Ranking spreaders by decomposing complex networks, Phys. Lett. A 377 (14) (2013) 1031–1035.
[7] J.G. Liu, Z.M. Ren, Q. Guo, Ranking the spreading influence in complex networks, Physica A (2013) http://dx.doi.org/10.1016/j.physa.2013.04.037.
[8] G. Sabidussi, The centrality index of a graph, Psychometrika 31 (4) (1966) 581–603.
[9] M.E.J. Newman, A measure of betweenness centrality based on random walks, Soc. Netw. 27 (1) (2005) 39–54.
[10] S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, K-core organization of complex networks, Phys. Rev. Lett. 96 (4) (2006) 040601.
[11] S. Carmi, S. Havlin, S. Kirkpatrick, S.Y. Havitt, E. Shir, A model of Internet topology using k-shell decomposition, Proc. Natl. Acad. Sci. 104 (27) (2007) 11150–11154.
[12] V. Batagelj, M. Zaversnik, Generalized cores, 2002, arXiv:cs.DS/0202039.
[13] V. Batagelj, M. Zaversnik, An $O(m)$ algorithm for cores decomposition of networks, 2002, arXiv:cs.DS/0310049.
[14] A. Garas, F. Schweitzer, S. Havlin, A k-shell decomposition method for weighted networks, New J. Phys. 14 (2012) 083030.
[15] A. Montresor, F. De Pellegrini, D. Miorandi, Distributed k-core decomposition, IEEE Trans. Parallel Distrib. Syst. 24 (2) (2013) 288–300.
[16] A.E. Sarıyuce, B. Gedik, G. Jacques-Silva, K.L. Wu, U.V. Catayurek, Streaming algorithms for k-core decomposition, Proc. of the VLDB Endowment 6 (5) (2013).
[17] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (14) (2001) 3200–3203.
[18] M.E.J. Newman, Spread of epidemic disease on networks, Phys. Rev. E 66 (1) (2002) 016128.
[19] R. Albert, H. Jeong, A.L. Barabási, Error and attack tolerance of complex networks, Nature 406 (6794) (2000) 378–382.
[20] C. Castellano, R. Pastor-Satorras, Thresholds for epidemic spreading in networks, Phys. Rev. Lett. 105 (21) (2010) 218701.
[21] M.G. Kendall, The treatment of ties in ranking problems, Biometrika 33 (3) (1945) 239–251.
[22] W.R. Knight, A computer method for calculating Kendall's tau with ungrouped data, J. Amer. Statist. Assoc. 61 (314) (1966) 436–439.

[23] W.W. Zachary, An information flow model for conflict and fission in small groups, J. Anthropol. Res. 33 (1977) 452–473.
[24] D. Lusseau, K. Schneider, O.J. Boisseau, P. Haase, E. Slooten, S.M. Dawson, The bottlenose Dolphin community of doubtful sound features a large proportion of long-lasting associations, Behav. Ecol. Sociobiol. 54 (2003) 396–405.
[25] P. Gleiser, L. Danon, Community structure in Jazz, Adv. Complex Syst. 6 (4) (2003) 565.
[26] J. Duch, A. Arenas, Community identification using extremal optimization, Phys. Rev. E 72 (2005) 027104.
[27] M.E.J. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (3) (2006) 036104.
[28] 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.
[29] N. Xie, Social network analysis of blogs, M.Sc. Dissertation, University of Bristol, 2006.
[30] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (1998) 440–442.
[31] M. Boguña, R. Pastor-Satorras, A. Diaz-Guilera, A. Arenas, Models of social networks based on social distance attachment, Phys. Rev. E 70 (2004) 056122.
[32] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: densification and shrinking diameters, ACM Tran. Knowl. Doscov. Data (ACM TKDD) 1 (1) (2007).
[33] J. Leskovec, K. Lang, A. Dasgupta, M. Mahoney, Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters, Internet Math. 6 (1) (2009) 29–123.
[34] A.L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[35] R. Albert, A.L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (1) (2002) 47.
[36] A. Lancichinetti, S. Fortunato, F. Radicchi, Benchmark graphs for testing community detection algorithms, Phys. Rev. E 78 (2008) 046110.
[37] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (2010) 75–174.