Journal of Informetrics 7 (2013) 223–239

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

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

C-index: A weighted network node centrality measure for collaboration competence

Xiangbin Yan[a,*], Li Zhai[a,b], Weiguo Fan[a,c]

[a] School of Management, Harbin Institute of Technology, Harbin 150001, China
[b] School of Applied Science, Harbin University of Science and Technology, Harbin 150080, China
[c] Pamplin College of Business, Virginia Tech, Blacksburg, VA 24061, United States

ARTICLE INFO

Article history:
Received 23 July 2012
Received in revised form 7 November 2012
Accepted 8 November 2012

Keywords:
Centrality measures
Scale-free networks
H-index
Weighted networks
Node strength

ABSTRACT

This paper proposes a new node centrality measurement index (c-index) and its derivative indexes (iterative c-index and c_g-index) to measure the collaboration competence of a node in a weighted network. We prove that c-index observe the power law distribution in the weighted scale-free network. A case study of a very large scientific collaboration network indicates that the indexes proposed in this paper are different from other common centrality measures (degree centrality, betweenness centrality, closeness centrality, eigenvector centrality and node strength) and other h-type indexes (lobby-index, w-lobby index and h-degree). The c-index and its derivative indexes proposed in this paper comprehensively utilize the amount of nodes' neighbors, link strengths and centrality information of neighbor nodes to measure the centrality of a node, composing a new unique centrality measure for collaborative competency.

1. Introduction

The research on scientific network as a typical informetric network has received more attention recently, such as citation and co-citation networks (Chen & Redner, 2010; Ding, Yan, Frazho, & Caverlee, 2009; Egghe & Rousseau, 2002), co-author networks (Liu & Shih, 2005; Newman, 2001a, 2001b; Rogriguez & Pepe, 2008), keywords networks (Su & Lee, 2010), and patent networks (Liu & Shih, 2005). As a major research form of scientific networks, the scientific collaboration network studies the partnership between the subjects of scientific research. For instance, the weighted co-author network refers to author pairs in co-author networks characterized by the number of times they collaborate (Zhao, Rousseau, & Ye, 2011). Complex relations in networks are reflected by different numbers of links and weights. In a social network, the research on the roles, positions, influence and centrality of nodes is usually expressed by node degree centrality (Freeman, 1979; Nieminen, 1974; Shaw, 1954), node strength (Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004), closeness centrality (Bavelas, 1950; Sabidussi, 1966), betweenness centrality (Freeman, 1979) and eigenvector centrality (Bonacich, 1972). Nonetheless they cannot accurately describe the collaboration competence of nodes.

How to measure the cooperation competence of nodes through the collaboration network in a weighted co-author network? In real life, if a person has many collaborators in close contact, most of whom are VIPs (of high social rankings and large influence), then his personal collaboration competence is strong. This thought can also be reflected in the weighted collaboration network. If a node has many neighbors whose linking edges are of great strengths and the neighbor nodes bear tremendous significance (or centrality, position, etc.), then the collaboration competence of this node is believed to

* Corresponding author. Tel.: +86 451 86414022.
E-mail addresses: xbyan@hit.edu.cn, hityxb@gmail.com (X. Yan).

1751-1577/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2012.11.004

be powerful. In essence, the collaboration competence of a node is related with node degree (amount of co-authors), edge strengths (collaboration frequency), and importance of neighbor nodes (co-authors). How to comprehensively consider these factors? It is an interesting thought to combine h-index (Hirsch, 2005) with the social network. H-index proposed by Hirsch (2005) is a measure integrating the amount and citation amount of papers to measure the academic achievements or influence of scholars. "A scientist has index h if h of his or her N_p papers have at least h citations each and the other $(N_p - h)$ papers have $\leq h$ citations each" (Hirsch, 2005). H-index simply and effectively considers the amount and citation amount of papers. It measures the key part of a dataset in a relatively natural way (Zhao et al., 2011). Since its introduction, the h-index and some related bibliometric indices have received a lot of attention from the scientific community in the last few years (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009).

Based on the idea of h-index, this paper proposes a collaboration index (hereinafter called c-index) to measure node centrality reflecting its collaboration competence in the weighted undirected network. It comprehensively encompasses node degree, node edge strength and collaboration competence of neighboring nodes. We define c-index of a node in a weighted network as the maximum integer c that makes the product of the edge strength of the node and the node strength of corresponding neighboring nodes no less than c (here the sum of edge strength of a node is taken as the strength of this node). Obviously, in a weighted collaboration network, if a scholar has many collaborators with whom he/she has frequent collaborations or some of the collaborators have strong competence in collaboration (here the strong competence in collaboration refers to the total frequency of a scholar to collaborate with others, namely the node strength of the network), then his/her c-index will be higher.

Some scholars have applied h-index in the social network and proposed some significant indexes. Zhao et al. stated that the h-degree in the weighted network can be used as the centrality measure of nodes. "The h-degree ($d_h(x)$) of node x in a weighted network is equal to k if k is the largest natural number such that x has at least k links each with strength at least equal to k" (Zhao et al., 2011). Based on the definition of the h-degree, Zhao and Ye (2012) introduced the directed h-degree and derived quantities for characterizing directed weighted networks. H-degree merely considers node degree and edge strength without considering the influence of the neighboring nodes. Nonetheless, the importance of neighboring nodes is also a vital factor influencing the centrality of a node, which is central to citation rankings, and web page ranking (e.g., PageRank used by Google). Many classic node centrality measurements have been established according to this factor, such as Katz Prestige (Katz, 1953) and eigenvector centrality (Bonacich, 1972). Schubert (2012a) proposed the partnership ability index, denoted as φ, and pointed out that his idea fits into the framework introduced in (Zhao et al., 2011). "An actor is said to have a partnership ability index φ, if with φ of his/her n partners had at least φ joint actions each, and with the other $(n - \varphi)$ partners had no more than φ joint actions each" (Schubert, 2012a). Partnership ability index φ is a special case of the h-degree (Rousseau, 2012). However, φ did not consider the influence of the importance of neighboring nodes upon the competence of nodes in collaboration. Furthermore, Schubert (2012b) analyzed the collaboration of Jazz musicians, the partnership ability index (φ) was found to be a useful measure to characterize the way performers are embedded in their partnership network. In addition, Schubert, Korn, and Telcs (2009) defined the network's h-index, which is the entire measure of the network applicable in comparison between different networks but inapplicable in measuring the characteristics of nodes. Korn, Schubert, and Telcs (2009) put forward the lobby-index (or l-index) for the non-weighted network to describe efficient communication. "The l-index or lobby index ($l(x)$) of a node x is the largest integer k such that x has at least k neighbors with a degree of at least k". Since the weighted network can be regarded as the non-weighted network, that is, ignoring the edge strength while only considering the existence of edge, the node in the weighted network can also be used to calculate l-index. Of course, there is a loss of information to ignore the influence of different edge strengths in calculating l-index of nodes in the weighted network. Zhao et al. (2011) promoted l-index to be the w-lobby index in the weighted network, "the w-lobby (weighted network lobby index, hereinafter called wl-index) index ($\mathrm{wl}(x)$) of a node x is the largest integer k such that x has at least k neighbors with node strength at least k". Campiteli, Holanda, Soles, Soares, and Kinouchi (2010) studied the nature of l-index combining the actual non-weighted network. The l-index is compared with the degree centrality, the betweenness and eigenvector centralities in the case of a biological network and a linguistic network. Efficient communication means high impact (wide access or high reach) and low cost (Korn et al., 2009). The l-index and wl-index are put forth as measures for effective communications, considering degree (strength) of a node or neighboring nodes and measuring the competence of effective communications through degree of nodes as neighboring nodes without considering the important factor that different edge strength have different communication competence.

Our proposed c-index measures the collaboration competence of a node in a weighted network by factoring not only the number of collaborators and the collaboration frequency, but also the collaboration competence of collaborators themselves. The idea of c-index is similar to h-index. However, it effectively measures the core of the above information and strike the balance of many sources using the product of the number of collaborators, the competence in collaboration, and the collaboration frequency. Its utility in measuring the collaboration competence of a node in a weighted network cannot be substituted with other indexes. Our comparison of c-index with other well-known bibliometric indexes reveals that it is apparently superior to the degree centrality (simple calculation of the number of collaborators), node strength (simple calculation of the sum of collaboration frequencies), h-degree (without utilizing the importance of the collaborators), l-index (without utilizing collaboration frequency) and wl-index (calculating the strength of neighboring nodes only through collaboration frequency). However the closeness centrality, the betweenness centrality and the eigenvector centrality are not directly comparable since they use different information in computation. Our later analysis of correlations among major

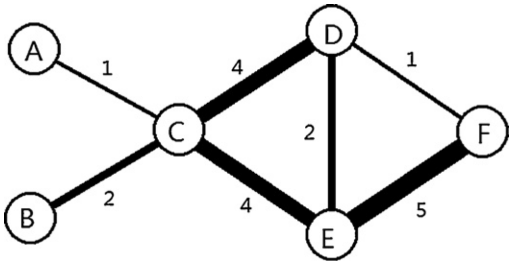

Fig. 1. Example of the c-index in a weighted network.

indexes indicates that the correlation between these three indexes, c-index and other derivative indexes is not too high, which confirms the uniqueness of our proposed c-index.

Based on the newly proposed c-index, this paper also proposes two significant derivative indexes, namely iterative c-index and c_g-index. The c-index and its derivative indexes proposed in this paper are not merely applicable in describing the collaboration competence in the collaboration network but also applied as a centrality measure in other networks; for instance, measuring the influence of a person in the network of friends and the competence of information spreading of nodes in the communication network, etc. The structure of the paper is as follows: Section 2 gives the definition of c-index and discusses its theoretical properties. It also discusses two derivative indexes of c-index. Section 3 conducts a case study to comprehensively evaluate the c-index and its derivative indexes in a large scientific collaboration network, along with other well-known centrality indexes. Section 4 concludes the paper and point out future directions.

2. Methodology

2.1. The c-index

Based on the factors influencing collaboration competence, this paper combines the characteristics of h-index and proposes the new collaboration index – c-index.

Definition 1 *(c-index).* The collaboration index $c(x)$ of node x is the biggest integer c such that the node x has at least c neighboring nodes satisfying that the product of each node strength and the strength of the edge linked with node x is no less than c.

In other words, the c-index of node x is the “h-index” of the sequence of the products of the strength of each neighbor node and the strength of each edge between the corresponding neighbor node and node x. Here we use “h-index” mainly as an operator, which is not the academic achievement measurement index but the calculation method reflected in the definition of h-index.

An operable definition is that arranging the product of the strength of each edge of node x and the node strength of each corresponding neighboring node, marking the same product with different serial numbers, when and only when the product of c before the arrangement is at least c and the product of the $c+1$th is less than $c+1$, the c-index of this node is c. If all the product values are larger than the node degree d, then the c-index of node x is the node degree d.

Obviously, c-index is the monotone nondecreasing function of node strength, edge strength and node degree. Consistent with Zhao et al. (2011) and Barrat et al. (2004), in this article we define the node strength of a node in a weighted network as the sum of the strengths (or weights) of all the links. For simplicity we usually assume that weights are natural numbers. For instance, edge strengths (or weights) in the co-authorship network are the collaboration times of two nodes.

Next we will illustrate the calculation of c-index. Fig. 1 gives an example of a weighted network and the line width indicates the size of strength with the specific numerical values shown in the figure. Supposing Fig. 1 is a co-authorship network, we calculate the c-index of Scholar C. Scholar C has four collaborators, namely A, B, D and E, who have respectively collaborated once, twice, four times and four times with C. The node strengths of A, B, D and E are 1, 2, 7(=4+2+1) and 11(=4+2+5). Then calculate the products of node strengths and their collaboration frequencies with C respectively, and rank the products in the descending order (shown in Fig. 2). It can be seen that C has at least 3 collaborators whose products are no less than 3. So c-index of C is 3.

Korn et al. (2009) point out that “Efficient communication means high impact (wide access or high reach) and low cost”. C-index can also measure the effectiveness of communications. Higher edge strengths indicate more frequent communications and the lower communication cost. Greater strengths of neighboring nodes indicate more powerful communication competence and greater influence. C-index comprehensively considers the two and calculates the product. The neighbor nodes can directly influence the communication competence of the node because of the neighbors of great influence. So c-index applies h-index to balance the above product and node degree. It is thus seen that c-index is suitable for describing the effective communication competence of nodes, which is consistent with the thought of collaboration competence.

Table 1
Degree, node strength, c-index, l-index, wl-index and h-degree of Fig. 1.

Index	Node					
	A	B	C	D	E	F
Degree	1	1	4	3	3	2
Strength	1	2	11	7	11	6
c-Index	1	1	3	3	3	2
l-Index	1	1	2	2	2	2
wl-Index	1	1	2	3	3	2
h-Degree	1	1	2	2	2	1

The node degree, node strength, c-index, l-index, wl-index and h-degree of the 6 nodes can be calculated according to Fig. 1 with the results shown in Table 1. The results indicate that both l-index and wl-index believe *C* and *F* have the same grade. Nevertheless Fig. 1 obviously shows that node *C* is more important than node *F* because they are linked with node *D* and *E*. Besides, *C* is also linked with *A* and *B*. Specific link strength also shows that *C* is more important than *F*. wl-Index is the unique method that treats *D* as more important than *C*, at least not inferior to *D* as shown in Fig. 1. Although h-degree also regards *C* as more important than *F*, it regards *F* as identical with *A* and *B* in terms of importance. This is not very logical and reasonable since *F* obviously has more neighbors and higher link strengths. Overall, we can see from this example that c-index more reasonably divides the results, namely: $A = B < F < C = D = E$.

Consider another example of a star network in Fig. 3, the edge strength *s* is the strength of node linked with node *A*. If the c-index of node *A* is *c*, then *c* is the biggest integer to make node *A* have *c* edges whose strengths are not less than $\sqrt{c}$.

Generally speaking, if we compare c-index with other indexes, we can have the following observations.

Proposition 1. *If the edge strengths of the weighted network are of natural numbers, then*

(1) *for any non-isolated node x*, $1 \le l(x) \le wl(x) \le c(x) \le d(x)$;
(2) *for any node x with degree 1 (x has only one collaborator),* $c(x) = d(x) = l(x) = wl(x) = 1$
(3) *for the globally coupled network which consists of M nodes,* $c(x) = d(x) = l(x) = wl(x) = M - 1$

where $d(x)$, $l(x)$, $wl(x)$, $c(x)$ and $d_h(x)$ are node degree, l-index, wl-index, c-index and h-degree of node *x*.

Proposition 1 (3) shows that if *M* scholars co-authored a paper, then c-index value of these *M* scholars are $M - 1$ in the weighted network of cooperation created by the paper where the edge strength is the number of cooperations. If there are

Serial number		Product
1	<	44 (=4×11)
2	<	28 (=4×7)
3	<	4 (=2×2)
4	>	1 (=1×1)

Fig. 2. Calculation of c-index of node *C* of Fig. 1.

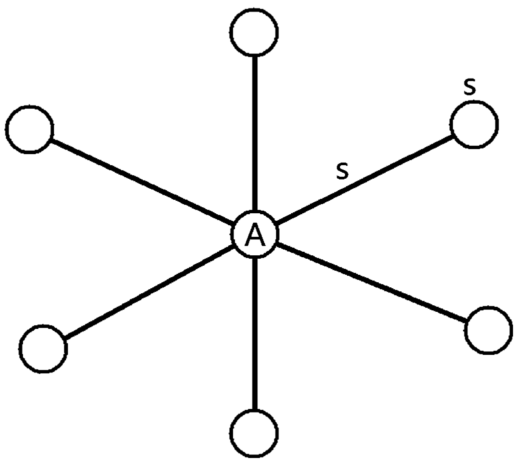

Fig. 3. Example of the c-index in a star network.

other cooperations in the M scholars but without other authors outside the M scholars, then the above conclusions still hold regardless of the conditions of cooperation.

More generally, we can make the following conclusions:

Proposition 2. *If edge strengths of the network are of non-negative real number, then the following will always hold*

1) *for any node x, $0 \le c(x) \le d(x)$*
2) *for any isolated node x, $c(x) = 1(x) = wl(x) = h_d(x) = 0$*

The weighted networks can also be regarded as the non-weighted networks, namely, neglecting the weight and only considering whether there exists a relationship or not between the nodes. Then:

Proposition 3. *Any node x in the non-weighted network permanently holds $l(x) = wl(x) = c(x)$, namely, the three are equivalent.*

The three conclusions are proved by the definitions of the indexes.

Next we will discuss the distribution of c-index, assuming the network is of scale-free weight, and edge strengths are of natural numbers.

Barabási and Albert (1999) proposed a very simple network model which has several key characteristics: most importantly the degree distribution has a power-law upper tail, the node degrees are independent, and typical nodes are close to each other. In what follows some properties of the c-index are investigated in weighted Scale Free (SF) networks. Barcza and Telcs (2009) prove the power laws in Scale Free networks distribution of the h-index. Let x be a randomly chosen author, $n(x)$ is the amount of paper of this author, y is a certain paper of this author and $cit(y)$ is the citation times of the paper y.

Note: in Theorem 1 and the following, b, b_1 and b_2 represent positive constants which can be different. To get the distribution of the c-index, the following lemmas should be used.

Lemma 1. *If the probability-distribution function of the random variable X is $f(x) \propto k^{-\gamma}$, then $\mathbb{P}(X \ge k) \propto k^{-(\gamma-1)}$, it is also valid otherwise* (Wang, Li, & Chen, 2006).

Theorem 1 (distribution of the h-index; Barcza & Telcs, 2009). *Assume that the productivity has an α-fat tail: $\mathbb{P}(n(x) \ge n) \approx bn^{-\alpha}$, and the citation score has a β-fat tail $\mathbb{P}(cit(y) \ge c) \approx bc^{-\beta}$, and assume that all the publication and citation scores are independent,*[1] *then $\mathbb{P}(h(x) \ge k) \cong k^{-\alpha(\beta+1)}$.*

Lemma 2. *If independently identically distributed random variables X and Y are positive integers, and $\mathbb{P}(X \ge k) \approx bk^{-\alpha}$, then $\mathbb{P}(X + Y \ge k) \approx bk^{-\alpha}$.*

Proof. By Lemma 1, it holds $P(X = k) \approx bk^{-(\alpha+1)}$, and with law of total probability we have

$$\mathbb{P}(X+Y=k) = \sum_{x=1}^{k-1}\mathbb{P}(X+Y=k,\ X=x) = \sum_{x=1}^{k-1}\mathbb{P}(X+Y=k|X=x)(X=x) = \sum_{x=1}^{k-1}\mathbb{P}(Y=k-x)\mathbb{P}(X=x)$$

$$\approx b\sum_{x=1}^{k-1}(k-x^{-(\alpha+1)}x^{-\alpha+1} \approx ck^{-(\alpha+1)})$$

Which together with Lemma 1 gives $\mathbb{P}(X + Y \ge k) \approx bk^{-\alpha}$

Next we will give the distribution of c-index.

Theorem 2 (distribution of c-index). *In the weighted undirected network G, the strengths of all edges are mutually independent. The strength of any edge S has distribution $\mathbb{P}(S \ge k) \approx bk^{-\beta}$ and all the node degrees are also mutually independent. The distribution of degree of any node x is $\mathbb{P}(d(x) \ge k) \approx bk^{-\alpha}$. When the strengths of all edges and degrees of all nodes are mutually independent, then $P(c(x) \ge k) \cong k^{-\alpha(\beta+3/2)}$.*

Proof. Calculate the distribution of the node strength first.

Suppose the number of the node in the network G is N, and S_x and $d(x)$ are the strength and degree of node x respectively. Clearly $\mathbb{P}(S_x = k, d(x) = n) = 0$ if $n > k$, then by full probability formula and Lemma 2, we deduce

$$\mathbb{P}(S_x=k) = \sum_{n=1}^{\min(N-1,k)}\mathbb{P}(S_x=k|d(x)=n)\mathbb{P}(d(x)=n) \approx b\sum_{n=1}^{\min(N-1,k)}k^{-(\beta+1)}n^{-\alpha} = bk^{-(\beta+1)}\sum_{n=1}^{\min(N-1,k)}n^{-\alpha} \approx bk^{-(\beta+1)}$$

thus $\mathbb{P}(S_x \ge k) \approx bk^{-\beta}$.

[1] Let $a_n \approx b_n$ mean that $a_n/b_n \to 1$ as $n \to \propto$ and $a_n \cong b_n$ that there is a $M > 1$ such that $1/M \le a_n/b_n \le M$ for all n.

Suppose S is the strength of any edge of node, and S_x is the strength of node x. Next we calculate the distribution of SS_x. Obviously, $S_x \geq S$. Similar with the proof of (1), it can be proved that $\mathbb{P}(S_x - S \geq k) \approx bk^{-\beta}$

$$\mathbb{P}(S_xS \geq k) = \sum_{y=1}^{+\infty}\mathbb{P}(S_xS \geq k, S = y) = \sum_{y=1}^{+\infty}\mathbb{P}\left(S_x \geq \frac{k}{y}|S = y\right)\mathbb{P}(S = y) = \sum_{y=1}^{+\infty}\mathbb{P}\left(S_x - S \geq \frac{K}{y} - y\right)\mathbb{P}(S = y)$$

$$= \sum_{y=1}^{\left[\sqrt{k}\right]-1}]\frac{k}{y} - y[^{-\beta}y^{-(\beta+1)} + \sum_{\left[\sqrt{k}\right]}^{+\infty}y^{-(\beta+1)} \doteq \sum_{y=1}^{\sqrt{k-1}}\left(\frac{k}{y} - y\right)^{-\beta}y^{-(\beta+1)} + \sum_{\sqrt{k}}^{+\infty}y^{-(\beta+1)} \approx b_1k^{-(\beta+1)/2} + b_2k^{-\beta} \approx bk^{-(\beta+1)/2}$$

Here operator $[x]$ is the biggest integer not more than x,$]x[$is the smallest integer nor less than x, "$x\doteq$" mean that Approximately equal.

(1) By (2), it holds $\mathbb{P}(S_x - S \geq k) \approx bk^{(-\beta+1)/2}$, which together with $\mathbb{P}(d(x) \geq k) \approx bk^{-\alpha}$ and Theorem 1 leads to $\mathbb{P}(c(x) \geq k) \cong k^{-\alpha(-\beta+3)/2}$.

Theorem 2 indicates that in the weighted scale free network (edge strengths are of natural numbers) if the strengths of all edges are independent of each other and have the index β in power laws, and the degrees of all nodes are also independent of each other and have the index α in power laws, and if the degrees and strengths of all nodes are also independent of each other, then c-index will have the index $\alpha(\beta + 3)/2$ in power laws.

2.2. Derivative indexes of c-index and their characteristics

In this section, we will define two derivative indexes related to the c-index proposed in the previous section. They are iterative c-index ($c^{(t)}$-index ($t \geq 2$) and their limit ic-index when $t \to +\infty$) and collaboration index based on g-index (c_g-index).

Since the original purpose of proposing c-index is to describe the collaboration competence of the node x, the construction of c-index integrates the contacts of the node x and its neighboring nodes, the amount of collaborators of the node x and the importance of the neighboring nodes. So c-index bears the characteristic of "self-referential" (Jackson, 2008), that is, the calculation of c-index is based on the premise that a node's importance is determined by how important its neighbors are. In order to more accurately measure the importance of the neighbor nodes in calculating the c-index, we can consider applying the method of c-index in the neighboring nodes and calculating their c-index as the measure of the importance and replacing the neighboring node strengths to calculate the c-index of node x, which can be called c-index of second order. Obviously, the information used in c-index of second order is enlarged from 2-step partial network (i.e., the network constituted of nodes that the distance to node x is not more than 2) to 3-step partial network of node x. Similarly, in order to obtain c-index of third order we take c-index of second order as the measure of the importance of the neighbor node to calculate the c-index of node x. We can further define c-index of fourth order, etc. We totally referred them as iterative c-index. Clearly, the iterative c-index of tth order of node x constantly modifies the importance of neighboring nodes through iterative c-index of $(t-1)$th order ($t \geq 2$). Next, we give the definition of iterative c-index.

Definition 2 *((iterative c-index)).* Note c-index as $c^{(1)}$-index. The $c^{(t)}$-index of node x noted as $c_t(x)(t \geq 2)$ is the biggest integer c_t that makes node x have at least c_t neighbor nodes satisfying that the product between $c^{(t-1)}$-index and the edge strengths of the neighbor nodes linked with node x should be no less than integer c_t, generally calling $c^{(t)}$-index ($t \geq 2$) as the iterative c-index. When $t \to +\infty$, limit $c^{(t)}(x)$ is the ic-index of node x, marked as $ic(x)$, namely, $ic(x) = \lim_{t\to+\infty} c^{(t)}(x)$.

We can notice that ic-index is completely "self-referential" (Jackson, 2008). When the $c^{(t)}(x)$ of all nodes in the network converges to limit $ic(x)$, it means that the $c^{(t)}(x)$ of all nodes in the network will not change as t increases. That is to say, the $ic(x)$ of any node x is the "h-index" of the sequencing of products of edge strengths of neighboring nodes and the value of ic-index of node x. Here, the ic-index of a node is decided by the ic-index of its neighboring nodes while the ic-index of the neighboring nodes is decided by their own neighboring ic-index (Jackson, 2008).

If node strength is noted as $c^{(0)}$-index, then the definitions of iterative c-index and c-index can be unified in Definition 2, noted as $c^{(t)}$-index ($t \geq 1$). Obviously, for any isolated node x in one network, it permanently holds $c_t(x) = ic(x) = 0(t \geq 1)$. As for the iterative c-index, we have the following theoretical properties:

Theorem 3. *For any given network G, the sequence $c^{(t)}(x),\ t \geq 1$ of iterative c-index of any node x is convergent.*

Proof. It is easy to know from the definitions of iterative c-index and h-index that the sequence $c^{(t)}(x),\ t \geq 1$ is monotone decreasing and $c^{(t)}(x),\ t \geq 0$ for all $t \geq 1$. So according to the monotone bounded convergence theorem, the order is convergent.

Theorem 3 indicates that ameliorating the measure of importance of neighboring nodes through constantly using the iterative c-index can reach a stable value which can be used as the significant measure of the collaboration competence of the node.

Table 2
ic-index and c_g-index of Fig. 1.

Index	Node					
	A	B	C	D	E	F
ic-Index	1	1	2	2	3	2
c_g-Index	3	4	8	8	9	7

Serial number		Product
1	<	12(=4×3)
2	<	12 (=4×3)
3	>	2(=2×1)
4	>	1 (=1×1)

Fig. 4. Calculation of $c^{(2)}$-index of node *C* of Fig. 1.

As an example, we calculated the iterative c-index and ic-index of nodes of the weighted network shown in Fig. 1. The results indicate that the iterative c-index is convergent in just one iteration, namely $c^{(2)}$-index is ic-index. The results of ic-index are summarized in Table 2.

We still use *C* as an example to demonstrate the calculation process of iterative c-index. Node *C* has four neighboring nodes, namely *A*, *B*, *D* and *E* whose contact strengths of node *C* are respectively 1, 2, 4 and 4. Table 1 manifests that the c-index of four neighboring nodes (namely $c^{(1)}$-index) are separately 1, 1, 3 and 3. Calculate their products and rank them in the descending order accordingly. Fig. 4 indicates that from the third item the sequence numbers are larger than the products. So the $c^{(2)}$-index of node *C* is 2. Similarly, the $c^{(2)}$-index of other nodes can be worked out. As we have the $c^{(2)}$=index of all nodes, the $c^{(3)}$-index of nodes can be calculated through repeating the above process. As for this case, the $c^{(3)}$-index of all nodes is equal to $c^{(2)}$-index. So for all the nodes, ic-index equals $c^{(2)}$-index.

Next we will discuss another derivative index c_g-index. This idea is sourced from an improved index g-index of h-index. Since h-index was proposed, its theoretical aspects have been thoroughly discussed (Barcza & Telcs, 2009; Egghe & Rousseau, 2006; Glänzel, 2006; Schubert & Glänzel, 2007; Ye, 2011). Through research on the advantages and disadvantages of h-index, many scholars put forth a series of derivative indexes (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2010; Batista, Campiteli, & Kinouchi, 2006; Bornmann, Mutz, & Daniel, 2008; Egghe, 2006; Jin, 2007; Kosmulski, 2006), among which g-index is the most widely applied. In analyzing the effects of h-index in 2006, Egghe proposed a g-index based on the past accumulative contributions. It breaks the limit of the bibliography total, which will be more fair for the scholars and institutions of few bibliography production but high citation frequency (Ding, Zhou, & Ye, 2008). Its definition is "A set of papers has a g-index *g* if *g* is the highest rank such that the top *g* papers have, together, at least g^2 citations. This also means that the top $(g+1)$ papers have less than $(g+1)^2$ papers"(Egghe, 2006). And g-index set when the accumulated citation times are always more than the square of serial number, no matter *g* value is less than or equal to the bibliography total, then it will increase by 1. If g-index is more than the bibliography total, the bibliography number increases by 1 and the accumulated citation does not change until the accumulated citation amount is less than the bibliography serial number square.

Since c-index is defined based on h-index, it has both the advantages and similar disadvantages of h-index. For instance in collaboration of scientific research, some scholars might have fewer co-authors but their collaboration strength is great or the co-authors are in important positions (Here it refers to a pivotal position in the collaboration network, such as strong competence in collaboration and a large amount of collaborators), making this node have powerful collaboration competence. But c-index based on h-index cannot reflect this. A simple case is that node *A* and node *B* in Fig. 1 have only one collaborator *C*. Despite different collaboration frequencies, the value of c-index is 1. Moreover, irrespective of whether the collaborators of node *A* and node *B* are of the same person, how important the collaborators are or how frequent they have collaborated, as long as node *A* and node *B* have only one collaborator, then their c-index is 1. For this reason, applying g-index as measure of the collaboration competence produces the following c_g-index which is more fair for scholars (nodes) having few co-authors (neighboring nodes) but high collaboration frequency (edge strength) or having important co-authors (neighboring nodes).

Definition 3. (c_g-index) node *x* has a c_g-index c_g if c_g is the highest rank such that the sum of the products of the edge strength of the top c_g node and the strength of corresponding neighbor node is at least ${c_g}^2$, noted by $c_g\,(x)$, namely $c_g(x) = c_g$.

Note: Here it is also set that c_g-index is not restricted by node degree. That is to say, when the accumulated product of strengths of edges and corresponding neighbor nodes is always larger than the square of serial number, no matter whether c_g is less than or equal to the node degree, it will increase by 1. When the c_g-index is larger than the node degree, then the serial number will increase by 1 but the accumulated product will not change until it is less than the serial number square.

Obviously, the c_g-index value of an isolated node *x* will be 0. Since $g \geq h$ (Egghe, 2006) the following conclusion will always hold.

Table 3
Calculation of c_g-index of node C of Fig. 1.

Serial number	Product (descending)	Square of serial number		Accumulative product
1	44	1	<	44
2	28	4	<	72
3	4	9	<	76
4	1	16	<	77
5		25	<	77
6		36	<	77
7		49	<	77
8		64	<	77
9		81	>	77

Proposition 4. *For any node x, it holds that* $c_g(x) \geq c(x)$.

As an example, we calculated the c_g-index of nodes of the weighted network shown in Fig. 1. The results of c_g-index are summarized in Table 2.

Table 3 manifests the calculation of c_g-index of node C. Rank the products obtained in Fig. 2 in the descending order, calculate the accumulated values of the first k $(k \geq 1)$ items and compare the accumulated values and that of k^2. Table 3 shows that until sequence number 8, the accumulated values of products of node C are always bigger than the square of the sequence number. Actually, $8^2 \leq 77 < 9^2$. So c_g-index of node C is 8.

Comparison between the results of Table 2 and Table 1 indicates that ic-index and c_g-index are different from other indexes. C_g-index points out that the grades of nodes are in $A < B < F < C = D < E$. The analysis of Fig. 1 shows that this grade ranking is reasonable and superior to other methods believing $A = B$, $E = D$ or C. Ic-index can also perfectly display the highest grade of E. Comparison with the c-index value indicates that using the c-index of the neighbor node to measure the collaboration competence and replacing the node strength will produce different results from the original c-index. This example displays the highest grade to recognize E and lowers the grades of C and D. The definitions of c-index, ic-index and c_g-index indicate the three indexes have different emphases. C-index is the most simple and natural in practice. Calculation of c-index of a node x utilizes the 2-step partial network centering on this node x. C_g-index also uses the 2-step partial network centering on this node, but it preferentially stresses the importance of edges with big strength and neighbor nodes with big node strength, which is fairer for the nodes of small degree and greater edge strengths or the neighbor nodes of powerful node strengths. Each iteration of $c^{(t)}$-index uses the value from previous time to measure the collaboration competence of the neighbor nodes, and subsequently modify the collaboration competence of the nodes until it is convergent. Using ic-index to describe the collaboration ability of the node emphasize more the influence of other nodes than c-index. The disadvantage of the ic-index is that it is more complicated than the first two in calculation. Since the iterative value is monotone decreasing, the ic-index values for a network with large-scale nodes but not in close contact might generate bad differentiation.

3. A case study

3.1. Data

We chooses nine top academic journals in the field of information systems to be the data source to construct co-author network, which are: European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of Information Technology, Journal of Management Information Systems, Journal of Strategic Information Systems, Journal of the Association for Information Systems, MIS Quarterly, Decision Support Systems. The data are retrieved from the Web of Science databases[2] on May 29, 2012, where the total 4759 articles (with 6496 authors) recorded by "Article" are downloaded from the above nine journals for the period of January 1, 1985–May 29, 2012.

Our previous discussion indicates that c-index and its derivative indexs (iterative c-index, ic-index, and c_g-index), wl-index and h-degree display different properties because of different edge strength measurements. For instance when the natural number is used as the edge strength, c-index and wl-index are equal to the node degree (see Proposition 1(3)). To check the differences between all indexes of different edge weights, we calculate the edge strength in two different ways to build two different weighted co-author networks.

3.2. Results

3.2.1. Network 1

Network 1 takes the scholar as the node, constructs the edge according to the collaboration relationship and makes collaboration frequency as the edge strength. That is to say, if two scholars have collaborated in $n (n \geq 1)$ papers, then they have linked edges and their edge strength is n. Otherwise they have no linked edge (Newman, 2001b).

[2] This paper accesses Web of Science databases through the library of Harbin Institute of Technology.

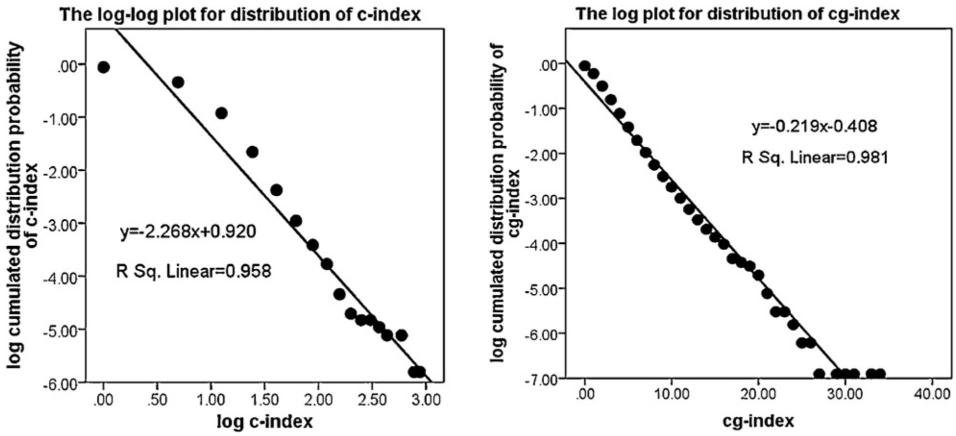

Fig. 5. The distribution of c-index and c_g-index.

Firstly we analyze the basic features of the co-author network and investigate the distribution of node degree (d), edge strength and node strength (s). After log–log transformation of cumulative probability distribution of all indexes, we observe that the distributions of obtained indexes all follow (approximately) power law distributions, so the co-author network can be regarded as a weighted scale free network (Barabási & Albert, 1999).

The average of the edge strength in the network is 1.129 (Here the average of the edge strength refers to the ratio of the sum of all edge strengths and the total of all edges in the network), that means the average number of cooperation between two co-authors is 1.129. Since the minimum value of edge strength in Network 1 is 1 (namely, only one collaboration), so the small scale of this number implies that most scholars in the network have only collaborated once. In reality, 90.8% of edge strengths of the network is 1. So cooperations within the network are often short-lived. Consequently the difference between all edge strengths is obscure. This means that the differences between the weighted network and the corresponding non-weighted network are also small.

Next, we calculate the c-index (c) and c_g-index (c_g) of all nodes in the co-author network, make log–log transformation of cumulated probability distribution of the c-index and log transformation of cumulative probability distribution of the c_g-index. Here cumulated probability distribution of the c-index adopts the following definition:

$$F(c) = Pr\{c - index \geq c\} = \frac{\#\{x : c(x) \geq c\}}{n},$$

where "#" means the number of elements in the set. The accumulated probability distribution of other indexes is defined identically. Regressing the transformed data and the distribution of obtained c-index is closely in the power laws, and the distribution of obtained c_g-index is closely in the exponential distribution. All the models and coefficients have all passed the significance testing with the results shown in Fig. 5.

Fig. 6 displays the sub-network of co-authors composed of 50 largest authors of c-index whose c-indexes are greater than or equal to 12. In the figure, the wider edge indicates the greater strength of linked nodes. The greater point shows the greater c-index. The marked number on the node is the c-index value of the corresponding node and the number on the edge shows the strength of the corresponding edge, namely the time of collaboration. It can be seen that even part of the network also displays that large authors of c-index tend to have more co-authors and greater collaboration strengths and exhibit the status of strong collaboration. Fig. 6 has no isolated scholars. They form three connected branches. That is to say, the collaborators of the scholars of high c-index often have scholars with high c-index. In the figure, the two big spherical groups are primarily obtained by two papers that have 20 authors and 17 authors respectively.

Next, we calculate the iterative c-index ($c^{(t)}$-index), the limit value ic-index of $c^{(t)}$-index of all nodes (as the $t \to +\infty$), and note c-index as $c^{(1)}$-index. The result shows that the $c^{(t)}$-index of all nodes is persistent when $t = 6$ and its convergence speed is fast. Fig. 7 is the scatter diagram of the absolute deviation of $c^{(t)}$-index and ic-index for the all 6496 nodes changing with t, where the absolute deviation is defined as $d_t = \sum_{x=1}^{N} |c^{(t)}(x) - ic(x)|$ $(t \geq 1)$, and N is the total number of nodes in the network. Fig. 7 shows that $c^{(2)}$-index or $c^{(3)}$-index can be used as a good approximation to ic-index. That is for the most of nodes the 3-step or 4-step partial network of this node includes all the collaboration information of this node. Fig. 8 shows the distribution of stable frequency of iterative c-index in the case of log–log transformation, namely the distribution of ic-index. It can be seen that ic-index obeys the power laws.

Next we calculate other h-type indexes of the nodes: l-index (l), wl-index (wl) and h-degree (d_h). To wholly examine the relationship between several indexes, we calculate their correlation. Since these indexes are all grade data, Spearman rank correlation is adopted here. See Table 4 for the results that all the indexes are in the positive relationship.

As the correlation of the degree and node strength reaches to 0.981, the differences between weighted network and the corresponding non-weighted network is not great. Except for the h-degree, the h-types indexes (l-index, wl-index, c- index

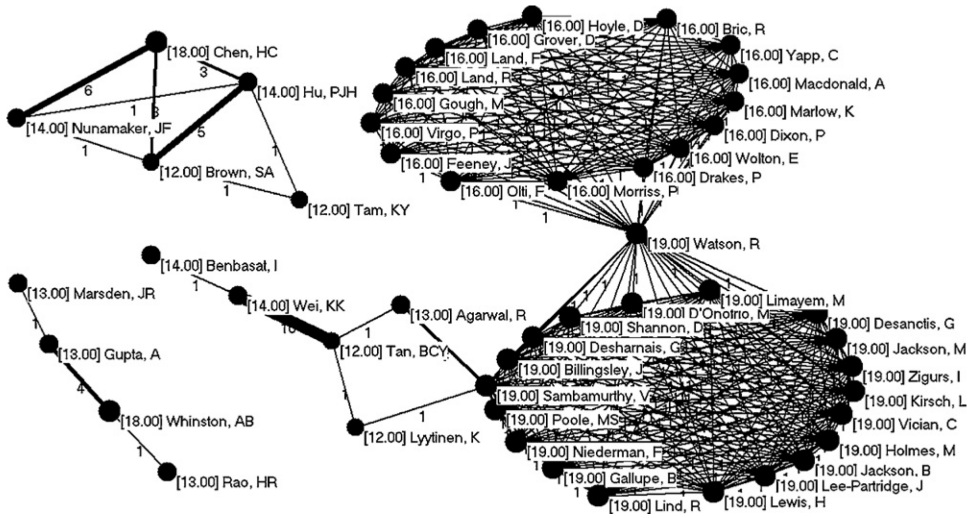

Fig. 6. Sub-network of co-authors (the 50 greatest c-index authors).

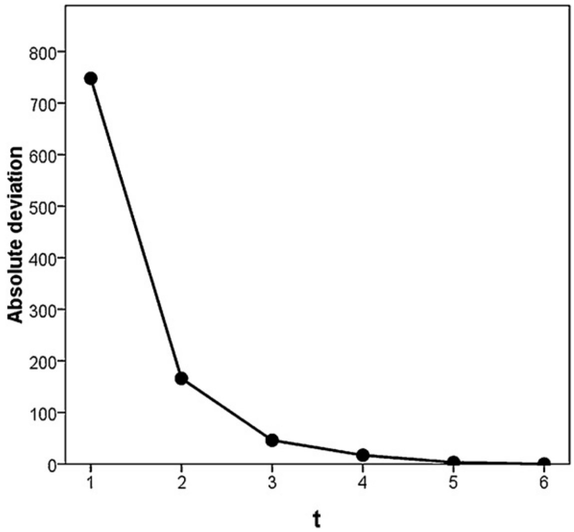

Fig. 7. Diagram of absolute deviation of $c^{(t)}$-index and ic-index changing with t.

Table 4
Spearman rank correlation of indexes in Network 1.

	d	s	cl	bw	ev	l	wl	dh	c	c_g	ic
Classic centrality measures											
d	–										
s	0.981**	–									
cl	0.620**	0.629**	–								
bw	0.624**	0.629**	0.476**	–							
ev	0.439**	0.451**	0.817**	0.399**	–						
Other h-type indexes											
l	0.948**	0.924**	0.594**	0.432**	0.420**	–					
wl	0.954**	0.931**	0.599**	0.450**	0.426**	0.996**	–				
dh	0.502**	0.554**	0.450**	0.375**	0.299**	0.469**	0.482**	–			
The indexes proposed in the paper											
c	0.955**	0.933**	0.601**	0.453**	0.426**	0.996**	0.999**	0.485**	–		
c_g	0.817**	0.840**	0.856**	0.511**	0.679**	0.808**	0.813**	0.532**	0.814**	–	
ic	0.941**	0.918**	0.583**	0.402**	0.410**	0.991**	0.991**	0.471**	0.991**	0.801**	–

** Correlation is significant at the 0.01 level (2-tailed).

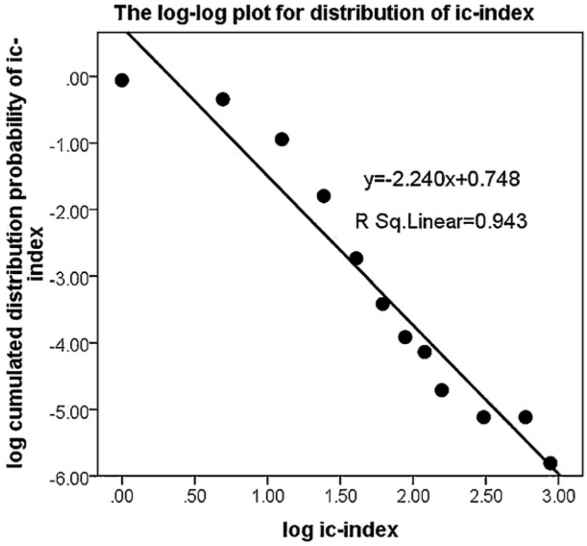

Fig. 8. Distribution of ic-index in the log–log coordinates.

c_g-index and ic-index) have a strong correlation with the node degree and node strength. This is because the edge strength is of natural number, and the average edge strength 1.129 of the network is close to the value 1 of the minimum edge strength 1, so these indexes of many nodes are the same, such as the isolated nodes and the nodes of degree 1 (see Proposition 1(2) and Proposition 2(2)). Proposition 1(3) demonstrates that if *M* scholars in the network construct the collaboration relationship through only one collaboration, then their c-index, wl-index and l-index are equal to the node degree. The prior analysis shows that most scholars in the collaboration network have collaborated only once, so many nodes correspond to the scenario that $c = l = wl = d$. Strongly related node degree and node strength constitute one major reason for high correlation of c-index, l-index, wl-index, node degree and node strength. L-index has comparative weak correlations with three classic centrality indexes (closeness centrality (cl), betweenness centrality (bw) and eigenvector centrality (ev)) and strong correlations with degree, which is consistent with the result of Korn et al. (2009).

C-index and ic-index have the greatest correlation with l-index and wl-index, For instance, c-index and wl-index hit 0.999, resulting from the characteristics of edge weight of the network. Since weighted collaboration network has no significant difference from corresponding non-weighted network, Proposition 3 shows that many nodes in the network will meet $c = wl = l$. In addition, it is obvious that if the edge strengths of nodes are all 1, hence $c = wl$ for this node. Besides, there are many isolated nodes with node degree = 1 in the network, hence $c = ic = l = wl$. All of these attribute to their high correlations. Additionally, compared with c-index, ic-index has lower correlation with all the other indexes, indicating that ic-index has greater differences from other indexes. Since ic-index is the limit of $c^{(t)}$-index which is defined on the basis of c-index, we hereby modify the importance of neighboring nodes through $c^{(t-1)}$-index($t \geq 2$). Consequently it is reasonable for ic-index and c-index to display strong correlation. Although c-index and ic-index have the greatest correlation with l-index and wl-index, both c-index and ic-index use the factor of edge strength which is not used by l-index or wl-index. Especially for the nodes of large edge strengths, c-index can more clearly manifest the collaboration competence of node than l-index and wl-index.

The performance of c_g-index is sharply different from that of c-index and ic-index. It has smaller correlation with node degree, l-index and wl-index. So c_g-index displays greater difference with other indexes for the reason that c_g-index changes the calculation of h-index into the method of g-index attaching more importance to high node strength and high edge strength on the basis of c-index, which is difficult for other indexes to display.

Propositions 1 and 2 show that, for the node with degree 1 or 0, the c-index and degree, l-index, wl-index are the same to each other. Therefore, in order to fully reflect the differences of the various indexes in a close-knit weighted network. We do the analysis by selecting the largest connected subgraph of the network of collaborators, and recursively removing the node with degree 1, after what, the network contains 3127 nodes.

Table 5 shows the results of the Spearman rank correlation of the indexes of the nodes in sub-network. Except the correlation of the betweenness centrality and the degree centrality, node strength and h-degree has increased, the other correlations have significantly decreased. Especially for the c_g-index, the largest correlation to all the other indexes is 0.721, which implies that the c_g-index has combined well the properties of the other centrality indexes but new to the others. Here *c* and *wl* still have the correlation of 0.999 with one major reason mentioned afore. There are many nodes whose edge strengths in the network are all 1, hence $c = wl$ for such nodes.

Although *c* has a comparatively higher correlation with *d*, *s*, *l* and *wl*, this does not affect its value of existence as the collaboration competence measurement index for the following reasons: Firstly, high correlation is influenced by the characteristics

Table 5
Spearman Rank Correlation of indexes in the subgraph of the Network 1.

	d	s	cl	bw	ev	l	wl	dh	c	c_g	ic
Classic centrality measures											
d	–										
s	0.952**	–									
cl	0.349**	0.353**	–								
bw	0.700**	0.710**	0.362**	–							
ev	0.288**	0.295**	0.505**	0.185**	–						
Other h-type indexes											
l	0.881**	0.818**	0.327**	0.404**	0.308**	–					
wl	0.897**	0.839**	0.335**	0.438**	0.302**	0.988**	–				
dh	0.359**	0.503**	0.231**	0.396**	0.205**	0.279**	0.315**	–			
The indexes proposed in the paper											
c	0.899**	0.843**	0.336**	0.444**	0.300**	0.986**	0.999**	0.325**	–		
c_g	0.664**	0.721**	0.678**	0.469**	0.509**	0.634**	0.648**	0.454**	0.649**	–	
ic	0.815**	0.758**	0.281**	0.332**	0.280**	0.926**	0.924**	0.283**	0.923**	0.584**	–

** Correlation is significant at the 0.01 level (2-tailed).

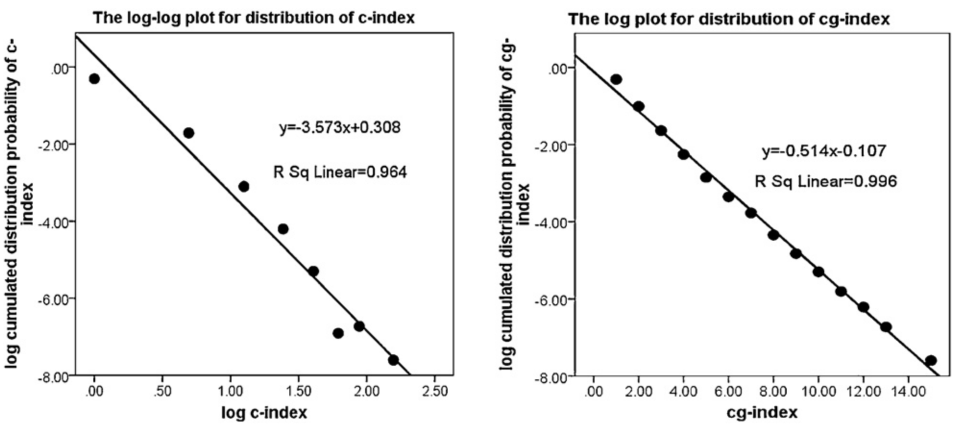

Fig. 9. The distribution of c-index and c_g-index.

of edge weight of this case; secondarily, unlike *l* and *wl*, c-index measures the centrality of node from the perspective of collaboration by giving more consideration to the influence of collaboration frequency degree upon the competence of collaboration. Compared with *d* and *s*, it considers the influence of importance of neighboring nodes with priorities.

3.2.2. Network 2

Using the number of cooperations to define the edge strength seems not very good. As mentioned earlier, 20 people working together on one paper, is far less in cooperating ability than that every two cooperating once alone. However, the current method of computing c-index using number of cooperations as the edge strength cannot differentiate that very well. Therefore, consistent with (Newman, 2001b), we use another weight calculation method to construct the co-author network. Suppose that scholars *A* and *B*, co-authored *M* papers, and we record the number of co-authors of each paper as $n_1, n_2, \ldots, n_M$. It is obvious that $n_i \geq 2, i = 1, 2, \ldots, M$. Then we define the edge strength S_{AB} between scholar *A* and scholar *B* as

$$S_{AB} = \sum_{i=1}^{M} \frac{1}{n_i - 1}$$

For example, three scholars *A*, *B* and *C* have only collaborated in writing one paper, then they are interlinked and the edge strength is 1/2. If *A* and *B* collaborate in one more paper, then the edge strength between *A* and *B* turns to 3/2 with other edge strengths unchanged. Obviously, here the node strength is the total number of papers in co-authorship with others published by the node (Newman, 2001b).

The analysis of network 2 is similar to network 1. An inspection of the node degree, edge strength and node strength shows that the network is a weighted scale-free network. The c-index distribution obeys approximately the power law distribution, and c_g-index obeys approximately the exponential distribution, and all the models and coefficients have all passed the significance testing with the results shown in Fig. 9. The results of the iterative c-index and the limit value of ic-index show that the $c^{(t)}$-index of all nodes is convergent when $t = 8$. Fig. 10 is the scatter diagram of the absolute deviation of $c^{(t)}$-index and ic-index of all the nodes changing with *t*. Fig. 11 shows the cumulative probability distribution of ic-index in the double logarithmic transformation, implying that ic-index obeys the power law distribution.

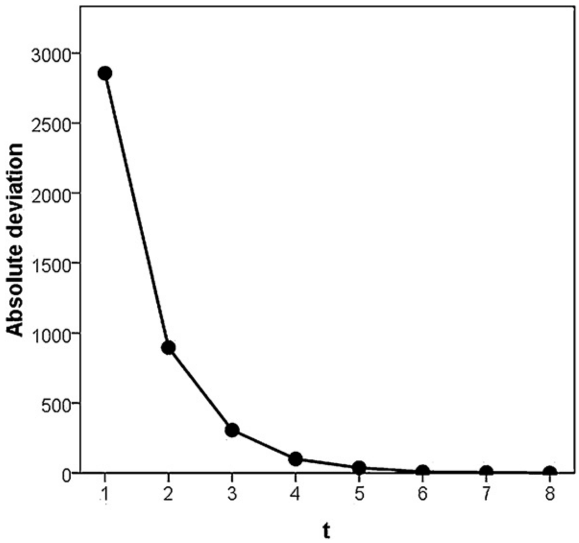

Fig. 10. Diagram of absolute deviation of $c^{(t)}$-index and ic-index changing with t.

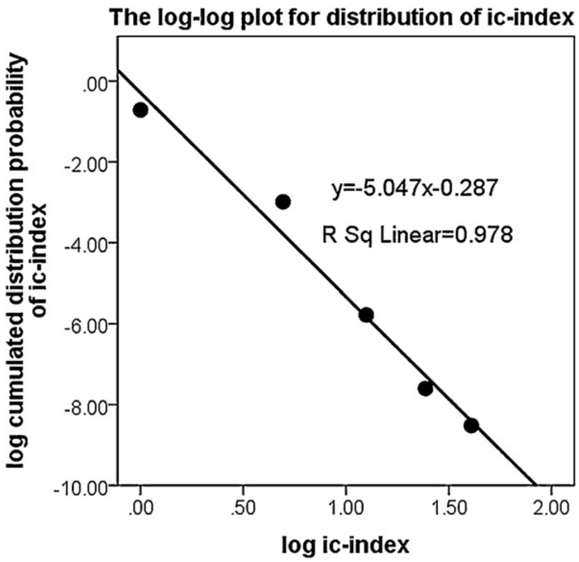

Fig. 11. Distribution of ic-index in the log–log coordinates.

Fig. 12 shows a co-authors sub-network by 32 scholars whose c-index are greater or equal to 5. The network is consisted by two large connected subgraph and two isolated nodes. Similar with Fig. 6, Fig. 12 also shows the collaborators of the scholars of high c-index often have scholars with high c-index. The biggest difference between Fig. 12 and Fig. 6 is that in Fig. 12 there are no fully coupled groups as in Fig. 6. In fact, in Network 2, publishing a paper by a large number of cooperation partners will not necessarily improve the c-index value, which is different from Network 1.

In order to visually observe the differences in various indexes, we arrange the authors according to the priority of the c-index, c_g-index and ic-index values in the descending order. Table 6 gives the index values of the first 32 nodes whose c-index are greater or equal to 5. It is worth noting that even though h-degree, l-index and wl-index presume the edge strengths are natural numbers in the original definition, their definitions still apply when the edge weights are real numbers. The last four columns in Table 6 shows the academic achievement measurement indexes of scholars, namely, *n* indicates the total number of published papers, *ct* represents the total citations, *h* refers to h-index and *g* means g-index. The calculation of these indexes is based on the data sets of this paper. From Table 6 it is clear to see that the ranking results according to different indexes of scholars are also different. This indicates that the node centrality measure proposed by us is different from the other indexes. How to understand their differences? Just as the analysis in Fig. 1, all indexes calculate the importance of nodes according to different principles. The advisable index is related to not merely the significance of the index itself but also the studied problem. It is known that degree centrality, node strength, l-index, wl-index and h-degree utilize less information than

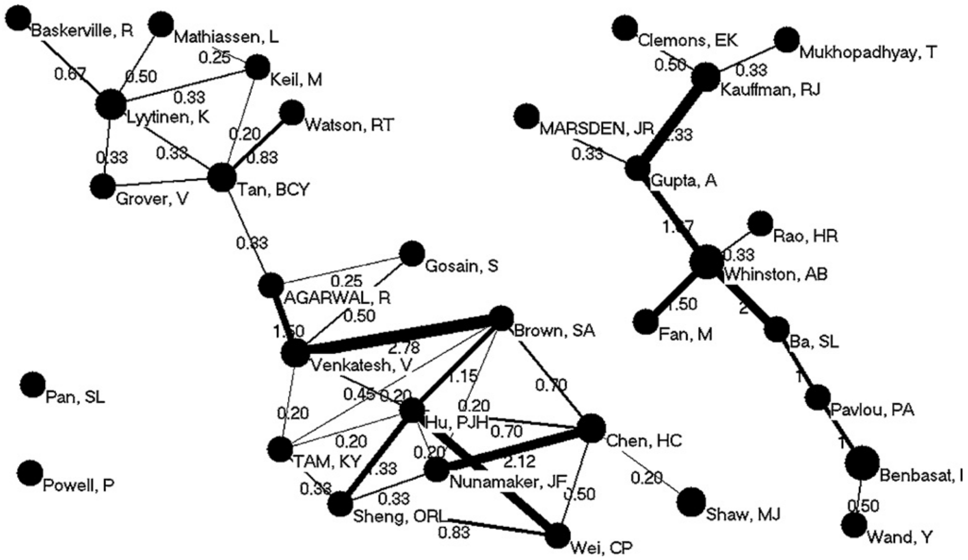

Fig. 12. Sub-network of co-authors (the 32 greatest c-index authors whose c-index are greater or equal to 5).

c-index, ic-index and c_g-index in considering the node centrality. C-index, ic-index and c_g-index comprehensively use the amount of neighboring nodes, edge strengths and centrality information of neighboring nodes to more accurately capture the collaborating capability of the focal node.

Observation of the four academic achievement indexes of 32 scholars in Table 6 shows that $n \geq 9$, $ct \geq 58$, $h \geq 3$, $g \geq 7$. In addition, the analysis of these four academic achievement indexes of all the 6496 scholars in the network unveils the

Table 6
Index values of the first 32 authors whose c-index are greater or equal to 5.[a]

AU	d	s	cl	bw	ev	c	ic	c_g	d_h	l	wl	n	ct	h	g
Whinston, AB	57	46	0.119	0.04	0.58	9	5	15	2	11	9	46	761	15	26
Benbasat, I	45	47	0.119	0.033	0.2	9	4	12	2	10	7	47	1947	22	44
Lyytinen, K	39	29	0.117	0.03	0.01	7	2	9	2	9	7	30	602	13	24
Tan, BCY	24	14	0.119	0.017	0.043	6	2	12	1	10	7	14	533	8	23
Kauffman, RJ	40	31	0.106	0.015	0.133	6	3	11	2	8	5	31	585	13	24
Chen, HC	80	40	0.106	0.021	0.003	6	2	11	2	11	8	40	660	13	24
Venkatesh, V	19	18	0.115	0.018	0.006	6	2	11	1	10	8	19	2169	9	46
SHAW, MJ	25	16	0.098	0.01	0.006	6	2	7	1	8	5	16	442	9	21
Gupta, A	39	26	0.111	0.015	0.258	5	4	15	2	10	6	26	316	10	17
Nunamaker, JF	47	22	0.113	0.028	0.003	5	2	12	1	10	8	22	316	10	17
Agarwal, R	35	26	0.12	0.017	0.018	5	3	11	2	10	6	26	1292	12	35
Ba, SL	11	8	0.113	0.008	0.19	5	3	11	1	7	6	9	783	7	27
Brown, SA	19	11	0.108	0.006	0.003	5	2	11	1	9	8	11	205	3	14
Fan, M	12	9	0.105	0.006	0.138	5	3	10	1	6	5	9	315	4	17
Pavlou, PA	16	13	0.116	0.012	0.06	5	3	10	2	8	6	14	993	7	31
Hu, PJH	26	14	0.11	0.009	0.002	5	2	10	1	10	8	14	58	4	7
Keil, M	37	21	0.116	0.024	0.005	5	2	9	1	8	6	21	745	12	27
Rao, HR	45	29	0.106	0.011	0.051	5	2	9	2	7	5	29	323	9	17
Wei, CP	21	14	0.101	0.003	0.001	5	2	9	1	7	5	14	93	7	9
Clemons, EK	13	13	0.102	0.004	0.031	5	2	8	1	7	6	18	280	9	16
Mathiassen, L	35	19	0.111	0.019	0.002	5	2	8	1	8	7	19	167	8	12
Tam, KY	29	21	0.11	0.011	0.002	5	2	8	2	9	7	23	832	10	28
Wand, Y	10	10	0.104	0.003	0.023	5	2	8	1	5	4	10	376	7	19
Watson, RT	24	16	0.109	0.007	0.017	5	2	8	1	8	5	16	326	10	18
Marsden, JR	20	16	0.106	0.006	0.033	5	3	7	1	9	6	16	76	5	8
Baskerville, R	17	12	0.108	0.005	0.002	5	2	7	1	7	5	13	252	6	15
Grover, V	32	25	0.113	0.017	0.012	5	2	7	2	7	5	26	857	12	29
Powell, P	17	14	0.087	0.007	0	5	2	7	1	7	5	14	255	9	15
Sheng, ORL	18	9	0.108	0.008	0.002	5	1	7	1	7	6	9	325	4	18
Gosain, S	15	9	0.108	0.003	0.006	5	2	6	1	5	5	10	231	7	15
Mukhopadhyay, T	14	11	0.099	0.003	0.008	5	2	6	1	6	5	11	106	6	10
Pan, SL	16	14	0.092	0.006	0	5	2	6	2	5	5	14	217	9	14

[a] The list is sorted based on the c-index value (c column).

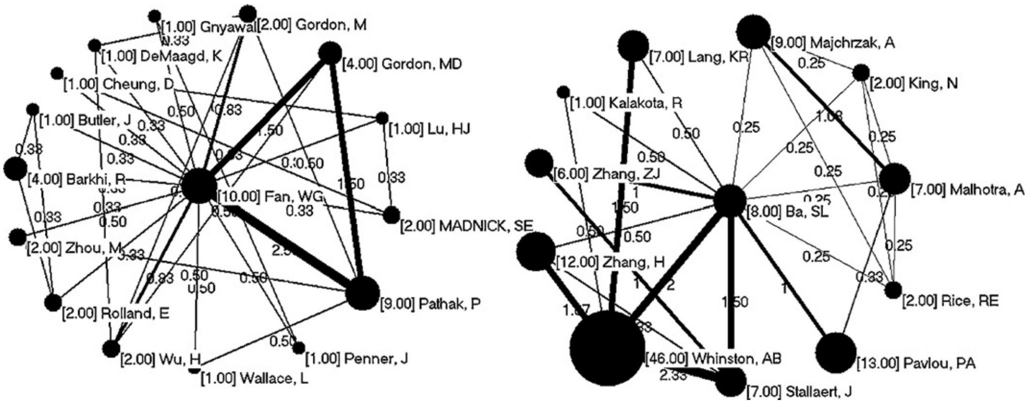

Fig. 13. The co-author networks of scholars Ba, SL and Fan, WG.

Table 7
Spearman rank correlation between c-index and other academic achievements.

	c-Index	c_g-Index	ic-Index	Number of papers	Citation	h-Index	g-Index
Number of papers	0.559**	0.576**	0.498**	–			
Citation	0.355**	0.365**	0.308**	0.516**	–		
h-Index	0.451**	0.468**	0.411**	0.756**	0.749**	–	
g-Index	0.356**	0.366**	0.307**	0.515**	0.992**	0.752**	–

** Correlation is significant at the 0.01 level (2-tailed).

following fact that for 1.8% of the scholars, $n \geq 9$; for 11.3%, $ct \geq 58$; for 10.0%, $h \geq 3$ and for 13.2%, $g \geq 7$. Thus, these selected 32 scholars of high c-index values also make great academic achievements to the information system fields.

The above results clearly show that famous scholars (i.e., those with four high academic achievement values in Table 6) in the information systems field tend to be listed as co-authors more often than non-famous ones. There could be multiple reasons behind this. One could be the resources and connections associated with high caliber researchers. Many of these famous IS scholars having many colleagues and students to collaborate with. Second reason could be that adding these scholars as coauthors increases the quality of the manuscripts and accordingly increase the chance of the paper acceptance in premier information system journals. Working with high caliber researchers in the corresponding discipline clearly is a trend being observed in many disciplines. The results, however, need to be interpreted with caution. The current case study only uses the 9 premier journals in the information systems field to construct the network. Had the journal list be expanded to include broad coverage of information science related journals such as JASIST and IP&M and other premier conference proceedings, the resulted indexes and rankings would have changed. A future task would be to increase the coverage of journals to a much broader list of publications to examine the differences in resulted rankings.

One can see from the results in Table 6 that high node strength value (s value) does not correspond to high c-index value (c value). In order to better understand this, we conducted another mini-case study. We picked one scholar from Table 6, Ba, SL, who is ranked in the 12th. We then picked another scholar Fan, WG, who is ranked in the 353th with the indexes are: $d = 15$, $s = 10$, $cl = 0.078$, $bw = 0.002$, $ev = 0$, $d_h = 1$, $l = 4$, $wl = 3$, $c = 2$, $c_g = 5$, $ic = 2$, $n = 10$, $ct = 90$, $h = 5$ and $g = 9$. In terms of both the number of co-authors and the number of cooperated publications, Fan, WG outperforms Ba, SL. However, his c-index, c_g-index and ic-index are smaller compared with Ba, SL. Fig. 13 shows The co-author networks of Ba, SL and Fan, WG. The labels for each node are node strength and scholar name, and the bigger diameter of the node implies stronger node strength. It can be seen from Fig. 13 that most of the edge strengths and the strength of most neighbor nodes of Ba, SL are bigger than the ones of Fan, WG. So overall Ba, SL has stronger collaboration ability.

Table 7 shows Spearman rank correlation between the three indexes (c-index, c_g-index, ic-index) and academic achievement (number of papers, citation, h-index and g-index). Academic achievements of scholars are computed based on the same 9 journal data set. The results suggest that there exists medium strength positive correlation between the three collaboration indexes and academic achievement indexes.

Finally, we calculate the Spearman rank correlation among the same set of indexes as in Table 4 and show the results on the Network 2 data in Table 8. If we compare the results from Tables 4 and 8, we can see that as the definition of edge strength of the network is changed, the correlations of indexes vary greatly. It can be seen from Table 8 that the indexes we proposed have low correlation with other indexes except for the 0.863 between ic-index and h-degree. Specific analysis of ic-index and h-degree of all nodes reveals that two indexes of 3313 (51.00%) nodes are concurrently 0 and those of 2393 (36.83%) nodes are contemporarily 1. Such phenomena result from the fact that the definitions of edge strengths in the Network 1 and Network 2 are not the same, and then the average edge strength of the Network 2 is somewhat low, which is 0.5303. It is easy to know from the definitions of h-degree and ic-index that small edge strengths exert the greatest influence upon these two indexes, which severely reduces the two index values of the nodes. However when the edge strengths of the

Table 8
Spearman rank correlation of indexes in Network 2.

	d	s	cl	bw	ev	l	wl	d_h	c	c_g	ic
Classic centrality measures											
d	–										
s	0.670**	–									
cl	0.620**	0.544**	–								
bw	0.624**	0.831**	0.477**	–							
ev	0.275**	0.358**	0.557**	0.364**	–						
Other h-type indexes											
l	0.948**	0.495**	0.594**	0.432**	0.249**	–					
wl	0.592**	0.643**	0.711**	0.562**	0.424**	0.542**	–				
dh	−0.152**	0.473**	0.015**	0.356**	0.137**	−0.291**	0.154**	–			
The indexes proposed in the paper											
c	0.273**	0.635**	0.581**	0.539**	0.424**	0.176**	0.709**	0.489**	–		
c_g	0.264**	0.644**	0.679**	0.527**	0.494**	0.164**	0.678**	0.486**	0.897**	–	
ic	−0.067**	0.534**	0.200**	0.421**	0.298**	−0.202**	0.321**	0.863**	0.634**	0.647**	–

** Correlation is significant at the 0.01 level (2-tailed).

network are non-negative real numbers, their minimum values are 0. Hence there exists the phenomenon of concentration in the direction of 0. This also explains the high correlation of h-degree and ic-index. Also for this reason, the correlation of the c-index and c_g-index increases a little than in Network 1. The correlation of h-degree and degree is negative, which is also because of the definition of the edge weights. Since majority of authors of our network only published one paper, more partners (namely greater node degree) means lower edge weights and smaller h-degree. Since l-index and degree are not affected by the edge weight of the network, their correlation is consistent with that in Network 1, to be 0.948. Overall, the several indexes we have proposed in this paper: the c-index, c_g-index, iterative c-index are shown to be unique and capture different nuances of the focal node centrality property.

4. Conclusions

Built upon some well-known scientific measurement index such as h-index, g-index, this paper proposes a new index (c-index) and its derivative indexes (iterative c-index and c_g-index) to measure the collaboration competence of a node in a weighted network. We prove that in the scale free weighted network, c-index obeys the power law distribution. A case study of a very large scientific collaboration network indicates that the indexes proposed in this paper are different from other common centrality measures (degree centrality, betweenness centrality, closeness centrality, eigenvector centrality and node strength) and other h-type indexes (l-index, wl-index and h-degree). The c-index and its derivative indexes proposed in this paper comprehensively utilize the amount of nodes' neighbors, link strengths and centrality information of neighbor nodes to measure the centrality of a node, producing more accurately utilized information and composing a new unique centrality measure for collaborative competency.

The collaboration competence of a node in the weighted network should be related with the amount of its collaborators (node degree), collaboration strengths (edge strengths) and collaboration competence or importance of collaborators (neighbor node strengths) and the collaboration competence should be the monotone non-decreasing function of the three. C-index is an index proposed to simply measure the collaboration competence of a node. The definition of c-index of a node should be the h-index of the product between the node edge strength and corresponding neighbor node strength. It uses the 2-step partial network of the node. $c^{(t)}$-index is to replace the neighbor node strength with the $c^{(t-1)}$-index of the node on the basis of the definition of c-index to measure the importance of the co-author. In this way the measure of the collaboration competence of node uses the $t+1$ step partial network of the node. We prove that the iterative c-index order is convergent and its limit value should be noted as ic-index. The example shows that the convergence of iterative c-index is quite fast and the disadvantage of ic-index is complicated calculation. Since the iterative values are monotone decreasing, the network ic-index value not in close connection might face bad differentiation. C_g-index as a derivative index of c-index mainly considers the significance of frequent collaboration with collaborators in important positions so as to apply g-index in the definition of collaboration index, which is fairer for the node of larger edge strength or higher neighbor node strengths. The c-index and its derivative indexes proposed in this paper measure the collaboration competence of nodes in the network more accurately and reasonably.

The three indexes proposed have different emphases. Since scientific evaluation does not advocate any single index, we can select the index in actual application by considering these emphases or make fair evaluation after comprehensively taking account of the value of each index. Although we discuss this type of collaboration indexes in accordance with the collaboration network, they can be used to indicate the importance, centrality and position of nodes in other complex networks as basic network node centrality measures.

Acknowledgements

This work is partly supported by the National Natural Science Foundation of PRC (nos. 71171067, 70890082) and Post-doctoral Science-research Developmental Foundation of Heilongjiang province (no. LBH-Q11114).

References

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). H-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*(4), 273–289.
Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2010). Hg-index: A new index to characterize the scientific output of researchers based on the h- and g-indices. *Scientometrics*, *82*(2), 391–400.
Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *5439*(286), 509–512.
Barcza, K., & Telcs, A. (2009). Paretian publication patterns imply paretian hirsch index. *Scientometrics*, *81*(2), 513–519.
Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of The United States of America*, *101*(11), 3747–3752.
Batista, P. D., Campiteli, M. G., & Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, *68*(1), 179–189.
Bavelas, A. (1950). Communication patterns in task oriented groups. *Journal of the Acoustical Society of America*, *22*, 725–730.
Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*(1), 113–120.
Bornmann, L., Mutz, R., & Daniel, H. D. (2008). Are there better indices for evaluation purposes than the h-index? A comparison of nine different variants of the h-index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, *59*(5), 830–837.
Campiteli, M. G., Holanda, A. J., Soles, P. R. C., Soares, L. H. D., & Kinouchi, O. (2010). *Hirsch index as a network centrality measure*. http://arxiv.org/abs/1005.4803
Chen, P., & Redner, S. (2010). Community structure of the physical review citation network. *Journal of Informetics*, *4*(3), 278–290.
Ding, N., Zhou, Y. B., & Ye, Y. (2008). The progress of the studies on h-index and h-type indices. *Document, Information & Knowledge*, (121), 72–77.
Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, *60*(11), 2229–2243.
Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, *69*(1), 131–152.
Egghe, L., & Rousseau, R. (2002). Co-citation, bibliographic coupling and a characterization of lattice citation networks. *Scientometrics*, *55*(3), 349–361.
Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, *69*(1), 121–129.
Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*(3), 215–239.
Glänzel, W. (2006). On the h-index—A methematical approach to a new measure of publication activity and citation impact. *Scientometrics*, *67*(2), 315–321.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of America*, *102*(46), 16569–16572.
Jackson, M. O. (2008). *Social and economic networks*. Princeton University Press.
Jin, B. H. (2007). The AR-index: Complementing the h-index. *ISSI Newsletter*, *3*(1), 6.
Katz, L. (1953). A new status index derived from sociometric analysis. *Psychomatrika*, *18*(1), 39–43.
Korn, A., Schubert, A., & Telcs, A. (2009). Lobby index in networks. *Physica A*, *388*(11), 2221–2226.
Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, *2*(3), 4–6.
Liu, D. R., & Shih, M. J. (2005). Hybrid-patent classification based on patent-network analysis. *Journal of the American Society for Information Science and Technology*, *62*(2), 246–256.
Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, *64*(1), 016131/016131–016131/016138.
Newman, M. E. J. (2001b). The structure of scientic collaboration networks. *Proceedings of The National Academy of Sciences*, *98*(2), 404–409.
Nieminen, J. (1974). On the centrality in a graph. *Scandinavian Journal of Psychology*, *15*(1), 332–336.
Rogriguez, M. A., & Pepe, A. (2008). On the relationship between the structural and socioacademic communities of a coauthorship network. *Journal of Informetrics*, *2*(3), 195–201.
Rousseau, R. (2012). Comments on A Hirsch-type index of co-author partnership ability. *Scientometrics*, *91*(1), 309–310.
Sabidussi, G. (1966). The centrality index of a graph. *Psychomatrika*, *31*(4), 581–603.
Schubert, A. (2012a). A Hirsch-type index of co-author partnership ability. *Scientometrics*, *91*(1), 303–308.
Schubert, A. (2012b). Jazz discometrics—A network approach. *Journal of Informetics*, *6*(4), 480–484.
Schubert, A., & Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, *1*(3), 179–184.
Schubert, A., Korn, A., & Telcs, A. (2009). Hirsch-type indices for characterizing networks. *Scientometrics*, *78*(2), 375–382.
Shaw, M. E. (1954). Group structure and the behavior of individuals in small groups. *Journal of Psychology: Interdisciplinary and Applied*, *38*(1), 139–149.
Su, H. N., & Lee, P. C. (2010). Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in technology foresight. *Scientometrics*, *85*(1), 65–79.
Wang, X. F., Li, X., & Chen, G. R. (2006). *The complex network theory and its application*. Tsinghua University Press.
Ye, F. Y. (2011). A unification of three models for the h-index. *Journal of the American Society for Information Science and Technology*, *62*(1), 205–207.
Zhao, S. X., Rousseau, R., & Ye, F. Y. (2011). H-Degree as a basic measure in weighted networks. *Journal of Informetrics*, *5*(4), 668–677.
Zhao, S. X., & Ye, F. Y. (2012). Exploring the directed h-degree in directed weighted networks. *Journal of Informetics*, *6*(4), 619–630.