

Analysis of an Optimal Measurement Index Based on the Complex Network

Hua Guan¹, Zhen Zhao², Lu Dai³

¹Information Engineering School, Hubei University of Chinese Medicine, Wuhan 430065, China

²Network and Educational Technology Center, Hubei University of Chinese Medicine, Wuhan 430065, China

³School of Computer, Dongguan University of Technology, Dongguan 523808, China
Email: dail@dgut.edu.cn

Abstract: In this paper, the discussion on the scientific cooperation network structure, the use of complex network analysis and social network analysis and the network analysis result from a scientific cooperation in both dynamic and static perspective. This paper extracts articles from 1995 to 2015 in conference proceedings, as experimental data sets with the corresponding network are called data mining cooperative network. The classic center-based index analysis was proposed by an improved node center metrics (*c-index*), weighted measure of the power of collaborative network node cooperation.

Keywords: Complex network, node centrality metrics (*c-index*), weighted cooperative network.

1. Introduction

Essentially, the cooperation strength and the degree of a node (the number of cooperators) and the edge's strength (the cooperation frequency) are related to the importance of the neighbor nodes (cooperators). So how should we consider these indexes? We shall improve the index by combining H-index with social network.

Firstly, we shall introduce the H-index. It is a method for measuring academic achievements or the scholars' influence by combining the number of citations with the number of articles, which was proposed by Hirsh in 2005 [1]. For instance, if one scientist owns NP (the number of articles) articles, and each of them has been cited for h times at least, while the sum of the times cited for other NP - h articles is less than h times, therefore, the scientist's H-index is h , which is demonstrated by the formula $[\max\{c|n(x) \geq c, d(y) \times S(y) \geq c\}]$, among which: $n(x)$ is the number of neighbor nodes of x ; y is any neighbor node of x ; $d(y)$ is the node degree of y ; $S(y)$ is the strength connecting the edges of x and y ; $[z]$ is the maximum integer that is not

bigger than z ; H-index measured the key parts of a dataset by a relatively natural method [2].

Based on the H-index, this paper proposed an improved cooperation index (hereinafter referred to as improved index) to measure the core strength of nodes and reflect the capacity of the nodes in the weighted undirected networks. It reflects on the nodes' degrees, strength and the cooperative ability of neighbor nodes. We define the index of nodes in the weighted undirected networks, as the strength of the nodes' edges and the nodes should be no less than the maximum integer of the nodes strength product of the neighbor nodes of c (defined the node strength as the sum of the strength of the nodes' edges). In one weighted cooperative network, if one scholar owns many cooperators, which are with high frequency, or these cooperators own strong cooperative ability (the strong cooperative ability here refers to the total frequency between the scholar and others, i.e., the nodes' strength in the network), so he/she owns high index [3]. H-degree only considered the nodes' degrees and the edges' strengths, while neglected the influence of the neighbor nodes. However, the importance degree of the neighbor nodes is a significant index influencing the core of the nodes, which is the core of citation ranking and webpage ranking.

2. Related work

The improved index measured the nodes' cooperative ability in the weighted networks, considered the number of the cooperators and the cooperation frequency, as well as the cooperators' cooperative ability. The index is similar to H-index, while it effectively measured the cores of the above information – the number of the cooperators, the cooperative ability and the cooperation frequency, as well as balanced the sources of all kinds of information. Its ability to measure the cooperative ability in the weighted networks cannot be replaced by other indexes [4, 5]. It can be found from the comparison with other famous literature metrologies that it is visibly more effective than the core of degree (the number of cooperators), the nodes' strength (the sum of the cooperation frequency), h-degree (without cooperators' importance degree and I-index (without cooperation frequency), and wl -index (calculate the strength of neighbor nodes only by cooperation frequency). However, due to the different information used for calculating, it is impossible to directly compare the closeness centrality with betweenness and eigenvectors concentration. In the following parts, we analyzed the relevance between key indexes, reflected that these three indexes are of low relevance and proved the uniqueness of index.

Based on the factors influencing the cooperative ability, this paper comprehensively summarized the characteristics of H-index, and proposed improved cooperation index, which measured the nodes' cooperative ability in the weighted networks as well as considered the number of cooperators and the cooperation frequency, as well as the cooperative ability of the cooperators themselves.

Definition 1 (c-index). The cooperation index $c(x)$ of the node x is the maximum integer c , the node x owns c neighbor node meets at least, sum of each node's strength and the node x 's strength connecting edges is not less than c .

The following parts demonstrated how to calculate the index c-index. Fig. 1 gave an instance for a weighted network. The width of the connecting lines referred to the strength, labeling with figures. Fig. 1 is a cooperator's network, and we calculated the index of scholar C. Scholar C owns four cooperators: A, B, D and E, who had been cooperated with C for 1, 2, 4 and 4 times. The node strength of A, B, D and E is 1, 2, 7(=4+2+1) and 11(=4+2+5). Then calculate the arithmetic product of each node's strength and the cooperation frequency, and ranking them in descending order (refer to Fig. 2). It is obvious that C owns 3 cooperators at least, thus the c-index for the scholar C is 3.

The greater the strength of the neighbor nodes, the greater the ability to communicate and the influence they will have. The index has taken both into consideration and multiplies them. Thus, the index adopts H-index to balance the arithmetic product and nodes' degrees. It follows from this that the index can be used to describe the effective communication ability between nodes, which is consistent with the cooperative ability.

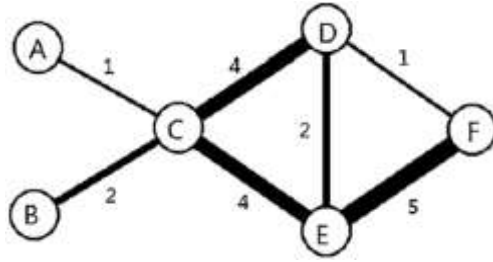


Fig. 1. Weighted artificial network for analyzing index

3. Example and conculation

In order to discuss the distribution of index, the following three lemmas shall be referred to.

Lemma 1. If the probability distribution function $f(x)$ of random variable X meets $f(x)ak^{-(\alpha-1)}$, thus $P(X \geq k) \approx ak^{-(\alpha-1)}$, which is also applicable to other cases [6].

Lemma 1 (distribution of index h) assumes one α -fat-tailed distribution: $P(n(x) \geq n) \approx bn^{-\alpha}$, citation score owns one fat-tailed $P(\text{cit}(y) \geq c) \approx bc^{-\beta}$, if all publishing scores are independent from citation scores, thus $P(h(x) \geq k) \approx k^{-\alpha(\beta+1)}$.

Lemma 2. If the independent identically distributed variables X and Y are positive integers, and $P(X \geq k) \approx bk^{-\alpha}$, thus $P(X+Y \geq k) \approx bk^{-\alpha}$,

$$\begin{aligned}
 (1) \quad P(X+Y=K) &= \sum_{X=1}^{K-1} P(X+Y=K, X=x) = \sum_{X=1}^{K-1} P(X+Y=k | X=x) = \\
 &= \sum_{x=1}^{K-1} P(Y=k-x)P(X=x) \approx b \sum_{x=1}^{K-1} (k-x)^{-\alpha} x^{-\alpha} \approx ck^{-(\alpha+1)}.
 \end{aligned}$$

Demonstration. According to the Lemma 1, $P(X = k) \approx bk^{-(\alpha+1)}$; and according to the total probability formula (1),

$$(2) \quad P(X + Y = K) \approx b \sum_{x=1}^{k-1} (k - x^{-(\alpha+1)}) x^{-\alpha+1} \approx ck^{-(\alpha+1)}.$$

According to the Lemma 2, $P(X+Y \geq k) \approx bk^{-\alpha}$, and it is followed:

Lemma 3 (distribution of index). In the weighted undirected network G , all edges' strengths are mutually independent, the weight distribution of some S edges is submitted to $P(S \geq k) \approx bk^{-\beta}$, and all the nodes' degrees are independent. The distribution of some nodes' degrees meets $P(d(x) \geq k) \approx bk^{-\alpha}$. When all edges' strengths are independent from the nodes' degrees, $P(c(x) \geq k) \cong k^{-\alpha(\beta+3/2)}$ [7].

Demonstration. Firstly, calculate the distribution of nodes' degrees. Assume the number of the network G 's nodes as N ; S_x and $d(x)$ are the node's strength and node's degree, respectively. If $n > k$, $P(S_x = k, d(x) = n) = 0$, it can be calculated according to the total probability formula and Lemma 2 that:

$$(3) \quad P(S_x = k) = \sum_{n=1}^{\min(N-1, k)} P(S_x = k | d(x) = n) P(d(x) = n) \approx \\ \approx b \sum_{n=1}^{\min(N-1, k)} k^{-(\beta+1)} n^{-\alpha} = bk^{-(\beta+1)} \sum_{n=1}^{\min(N-1, k)} n^{-\alpha} \approx bk^{-(\beta+1)}.$$

Therefore $P(S_x - S \geq k) \approx bk^{-\beta}$. Assume that S is the edges' strengths of some nodes, and S_x is the strength of node x . Now we'll calculate the distribution of $S_x - S$. Apparently, $S_x \geq S$. According to the demonstration of (1), we can obtain that $P(S_x - S \geq k) \approx bk^{-\beta}$,

$$(4) \quad P(S_x - S \geq k) = \sum_{y=1}^{+\infty} P(S_x - S \geq k, S = y) = \sum_{y=1}^{+\infty} P(S_x \geq \frac{k}{y} | S = y) P(S = y) = \\ = \sum_{y=1}^{+\infty} P\left(S_x - S \geq \frac{K}{y} - y\right) P(S = y) = \sum_{y=1}^{\lfloor \sqrt{k} \rfloor - 1} \left[\frac{k}{y} - y \right]^{-\beta} y^{-(\beta+1)} + \sum_{\sqrt{k}}^{+\infty} y^{-(\beta+1)} = \\ = \sum_{y=1}^{\sqrt{k-1}} \left(\frac{k}{y} - y \right)^{-\beta} y^{-(\beta+1)} + \sum_{\sqrt{k}}^{+\infty} y^{-(\beta+1)} \approx b_1 k^{-(\beta+1)/2} + b_2 k^{-\beta} \approx bk^{-(\beta+1)/2}.$$

The operator $[x]$ is the maximum integer that is no bigger than x , $\lceil x \rceil$ is the minimum integer that is no smaller than x , and " \approx " refers to approximately equal.

1) Formula (1) is according to formula (4), and $P(S_x - S \geq k) \approx bk^{-(\beta+1)/2}$ is established; 2) according to $P(d(x) \geq k) \approx bk^{-\alpha}$ and Lemma 1, $P(c(x) \geq k) \cong k^{-\alpha(\beta+3)/2}$ is established.

Lemma 2 demonstrated that within the weighted free networks (the edges' strengths are natural numbers), if all edges are mutually independent and meet the power law of β , all nodes' degrees are mutually independent and meet the power-law distribution of β , and all nodes' degrees and strengths are mutually independent, thus the index meets the power law of $\alpha(\beta+3)/2$ [8].

Therefore, we analyzed the basic characteristics of the cooperator's network and studied the node degree (d), edge strength and node strength (S). After the

LOG-LOG conversion of all indexes' cumulative probability distribution, we found that these indexes' distributions are approximately submitted to the power law. Therefore, Fig. 2 cooperator's network can be taken as a strengthened scale-free network [9, 10].

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

Fig. 2. The index calculation process for node C

Then we calculated all the nodes' index (c-index) in the cooperation network, and then had double logarithmic transformation on the cumulative probability distribution of c-index. “#” refers to the number of elements in the set. Other indexes of the cumulative probability distribution are defined similarly [11-13]. The distribution of c-indexes after transformed data regression approximates power-law distribution. All models and coefficients passed the tests, which is as shown in Fig. 3.

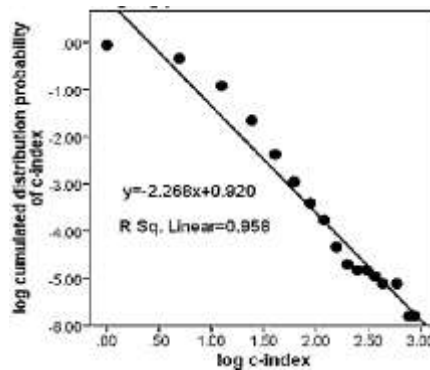


Fig. 3. Distribution of the index

Fig. 4 demonstrated the cooperator's subnet, including 50 authors of the maximum indexes. Their indexes are greater than or equal to 12. In the figure, the wider the edge is, the stronger the strength of the connecting node indicates; and the larger the node is, the higher the index is indicated. The index value is marked on the node, and the strength value of the edge is marked on the edge as well, i.e., cooperation duration. It follows that even if a part of the network is able to demonstrate the authors of high index, more cooperators and stronger cooperation strength and strong cooperation status.

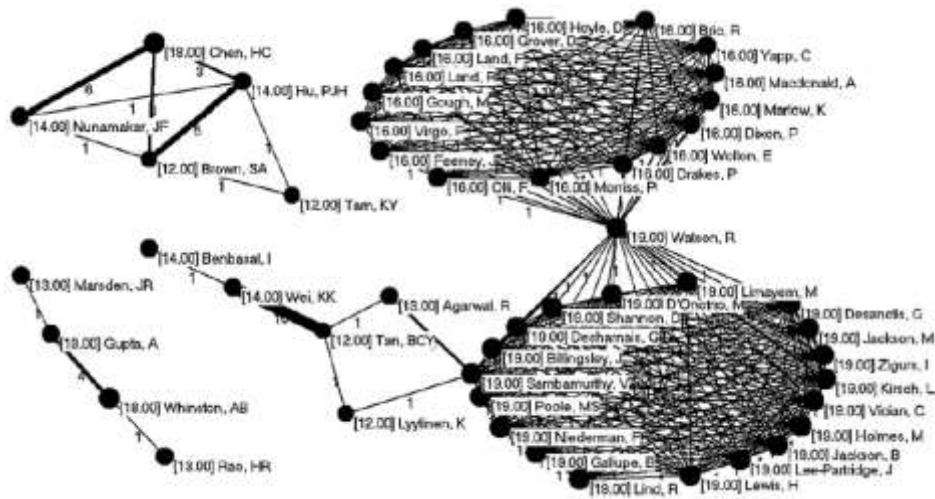


Fig. 4. 50 scholars according to the improved index

4. Analysis of improved index based on comparison to the other index

According to Definition 1, l-index is defined as the node center metrics of the node x and the node x 's strength connecting edges is not less than l , wl-index is defined as the node center metrics of the node x and the node x 's strength connecting edges is not less than wl , and H-degree (dh). The comprehensive inspection correlation of the relationship between these factors were calculated. Because of these factors is grade data, the Spearman rank correlation was used, the factors of the results are shown in Table 1.

Table 1. The analysis of Spearman correlation coefficient between several index

Index	c	dh	wl	l	ev	bw	cl	s	d
c	–	0.485	0.999	0.996	0.426	0.453	0.601	0.933	0.955
Other h-degree indexes									
dh		–	0.482	0.469	0.299	0.375	0.450	0.554	0.502
wl			–	0.996	0.426	0.450	0.599	0.931	0.954
l				–	0.420	0.432	0.594	0.924	0.948
The classic centrality metrics									
ev					–	0.399	0.817	0.451	0.439
bw						–	0.476	0.629	0.624
cl							–	0.629	0.620
s								–	0.981
d									–

In Table 1, because of the degree of relevance between node strength reached 0.981, so the weighted network and the corresponding unweighted network difference is not big. H-types except h-degree indexes (l-index, wl-index, c-index) and strongly correlated to the node degree, node strength. Since the edge strength for natural numbers, network of average intensity is close to the minimum while strength 1 value of 1, the index of many nodes (such as the node of isolated nodes and the degrees of 1) is same. If the network relationship of M scholars were

consisted only through one cooperative partnership, then their c -index, wl -index, l -index and the node degrees are equal. Many analysis show that most scholars cooperated only once, so many nodes in accordance with the following situations, $c = l = wl = d$. Strong related node degree and strength of the node is c -index, l -index, wl -index, node degree, node strength is the main reason for the high correlation. The index l -index is the third level center index close to the center (cl), middle center (bw), characteristic vector center degrees (ev) is a relatively weak correlation, and strongly correlated to the degree, this has to do with Korn and the other results [14]. The C -index, IC -index, l -index and wl -index has a maximum correlation. For example, c -index, wl – the index is 0.999, which is caused by the characteristics of the network edge right. Because the weighted network with no significant differences is corresponding to unweighted network, so the Lemma 3 shows that many nodes in the network will meet $c = wl = 1$. In addition, obviously, if the node of edge strength are all 1, then this node $c = wl$. Besides there are many isolated nodes, in the network, the node degrees = 1, so $c = IC = l = wl$. All these lead to the high correlation.

5. Conclusions

This paper put forward an improved factor for the centers of classic metric (center degrees, proximity, mediation, characteristic vector center) and other factors such as l -index, wl -index and h -degree. The index of comprehensive use the number of neighbor nodes, adjacency node connection strength and the concentration to measure the concentration, which provide more accurate information to the cooperate ability [15].

In the scale-free weighted networks, the index follows a power law, in the weighted network the cooperation ability of nodes is up to the number of neighbor nodes (nodes), cooperation intensity (intensity) and the importance of cooperate or collaborator (adjacency node connection), ability to cooperate for the three monotonous reduction function.

The improved index is used to simply node cooperation measurement. The definition of c -index should be defined as the product of strength and the adjacent point strength of the node x . Although the discussion of the type is suitable for cooperation network, the cooperation factor could be used also for other nodes in complex networks.

Acknowledgments: This work has been supported by the National 863 Program of China under Grant No 2013AA102302, Educational Research Program of Hubei University of Chinese Medicine under Grant No 2015A10, Research Program of Hubei Provincial Collaborative Innovation Center of Preventive Treatment by Acupuncture and Moxibustion Grant No HBPCIC-2016-011, Research Program of Department of Housing and Urban-Rual Development of Hubei Provincial under Grant No 2014-47 and the Popular science series into the campus of Guangdong Province (No 2014A070710018).

References

1. Litvak, N., W. R. W. Scheinhardt, Y. V. Volkovich. Probabilistic Relation between In-Degree and PageRank. – In: Proc. of 4th International Workshop WAW'2006, 2006, Banff, Canada, pp. 72-83.
2. Liu, L. G., Z. G. Xuan, Z. Y. Dang, Q. Guo, Z. T. Wang. Weighted Network Properties of Chinese Nature Science Basic Research. – Physica A-Statistical Mechanics and Its Applications, Vol. **377**, 2011, No 1, pp. 302-314.
3. Ma, N., J. Guan, Y. Zhao. Bringing PageRank to the Citation Analysis. – Information Processing and Management, Vol. **44**, 2008, pp. 800-810.
4. Milgram, S. The Small World Problem. – Psychology Today, Vol. **1**, 1967, No 2, pp. 60-67.
5. Nascimento, M. A., J. Sander, J. Pound. Analysis of SIGMOD's Coauthorship Graph. – SIGMOD Record, Vol. **32**, 2003, No 3, pp. 8-10.
6. Newman, M. E. J. Scientific Collaboration Networks: I. Network Construction and Fundamental Results. – Physical Review E, Vol. **64**, 2001, 016131.
7. Newman, M. E. J. The Structure of Scientific Collaboration Networks. – Proc. of the National Academy of Science of the United States of America, Vol. **98**, 2001, No 2, pp. 404-409.
8. Page, L., S. Brin. The Anatomy of a Large-Scale Hypertextual Web Search Engine. – Computer Networks and ISDN Systems, Vol. **30**, 1998, pp. 107-117.
9. Paullay, I. M., G. M. Alliger, E. F. Stoneromero. Construct Validation of 2 Instruments Designed to Measure Job Involvement and Work Centrality. – Journal of Applied Psychology Volume, Vol. **79**, 2003, No 2, pp. 224-228.
10. Perra, N., S. Fortunato. Spectral Centrality Measures in Complex Networks. – Physical Review E, Vol. **78**, 2008, 036107.
11. Pinski, G., F. Narin. Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics. – Information Processing and Management, Vol. **12**, 1976, pp. 297-312.
12. Price, J. D. S. Networks of Scientific Papers. – Science, Vol. **149**, 1965, pp. 510-515.
13. Newman, M. E. J. The Structure and Function of Complex Networks. – SIAM Review, Vol. **45**, 2003, No 2, pp. 167-256.
14. Newman, M. E. J. A Measure of Between Ness Centrality Based on Random Walks. – Social Networks, Vol. **27**, 2005, pp. 39-54.
15. Nisonger, T. E., C. H. Davis. The Perception of Library and Information Science Journals by LIS Education Deans and ARL Library Directors: A Replication of the Kohl-Davis Study. – College & Research Libraries, Vol. **66**, 2005, pp. 341-377.