Journal of Applied Science and Engineering Innovation Vol.2 No.4 2015 pp.102-105
ISSN (Print): 2331-9062
ISSN (Online): 2331-9070

The Application of Improve Author's Personal Influence of Network Ranking Algorithm

Jiatai Gang, Feiyue Liu, Zhe Zhang, Shujuan Li
College of Information Engineering, Dalian University, Dalian, 116622, China

Abstract: A new algorithm is provided in this paper based on LeadRank algorithm, which take Paul Erdos co-author network as an example with the author's personal influence ,the co-author's influence and the times of cooperation. The improved algorithm can comprehensive evaluate the author's influence in co-author network. And quantificational analyzing how to choose collaborator to promoting the academic influence of authors from three aspects: the co-author's influence, Clustering coefficient and the degree of node.

Keywords LeaderRank algorithm; Co-author network; Clustering coefficient; Academic influence

INTRODUCTION

With the booming development of academic research cooperation, co-author networks gradually become the hot point which scholars focus on. Co-author network refers to complex network of relationship between the co-author, and it's one of the social networks. Co-author network plays an important role in Entity Organization. Researchers often measure authors' influence through analysing co-author network, and make a corresponding decision to improve personal influence.

The essential of researching the co-author phenomenon is researching the relationship between authors. Social network analysis is known as studying relationship. Liu Xiaoming[1],et al analysed co-authership networks in the digital library through social network analysis; Zuo Meiyun[2], et al studied the influence of authors from the topological structure, and showed the individual states by analyzing network centrality; Linyuan Lü[3] applied LeaderRank algorithm to measure the influence of authors in co-author network, but this algorithm only considered influence of the author's co-authors; Li liang and Zhu qinghua[4] studied the co-author network from the angle of centrality analysis ,subgroup analysis and core-periphery analysis, they used simple binary network to indicate the co-author network based on ignoring some information of actual data. Zhang fuzeng[5],et al builded a directed network, qualitative presenting effectively measures and ways to promote the influences of scientists.

One of the most famous academic co-authors was the 20th-century mathematician Paul Erdös who had over 500 co-authors and published over 1400 technical research papers. Mathematicians always studied co-author network through analysis of Paul Erdös's co-author network in that era[6].This paper provided improved LeaderRank algorithm with H-factor, which take Paul Erdös co-author network as an example. The new algorithm comprehensive evaluate the author's influence from the author's personal influence, the co-author's influence and the times of cooperation in co-author network; and quantificational analyzing how to choose collaborator to promoting the academic influence of authors as soon as possible from three aspects: the co-author's influence, Clustering coefficient and the degree of node.

THE CONSTRUCTION OF NETWORK MODEL

The co-author network

Paul Erdös co-author network includes three co-authorship: Erdös with Erdös1, Erdös1 with Erdös1 and Erdös1 with Erdös2, as shown in Figure 1.

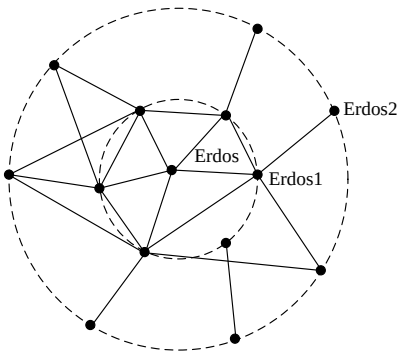

Figure.1. three co-authorships diagram

This paper constructs three cooperative relationships, only analyzing the influence of Erdös1. According to Figure 1, co-authors are numbered 1 to 511 in the order they appear in the file Erdös1.html. This paper Calculate the times of cooperation between Erdös1 and Erdös, and count the number of Erdös2, which had co-authered with Erdös1. Binary matrix of cooperation between Erdös1 would be builded. (If there are co-authorship between Erdös1 recorded as 1, if not, recorded as 0). So we can calculate the times of cooperation between Erdös1.

Corresponding Author: Jiatai Gang, College of Information Engineering, Dalian University, Da Lian 116622, China.

Improved LeaderRank algorithm

Leaderrank algorithm is a new effectively ranks web pages based on the hyperlink network, which was devised by Linyuan Lü,et al in 2011. The algorithm is a new algorithm improved based on PageRank. Its core idea is: introducing a ground node connects to every node through bidirectional links, thus the network becomes strongly connected.

Given node A connected with nodes $T_1, T_2, \cdots T_n$, take the influence value of node A as $PR(A)$, the influence value of $T_i (i = 1, 2, \cdots, n)$ is $PR(T_i)$, and the links of node T_i is $C(T_i)$. A "ground node" is introduced to LeaderRank algorithm, so adding 1 to the links of every node. The formula[7] of node A influence as follow:

$$PR(A) = \sum_{i=1}^{n+1} \frac{PR(T_i)}{C(T_i)} \quad (1)$$

In this paper, Erdös1 network is a network topology, each author as a node, Paul Erdös as the initial node—Ground in the author network. Using the Leaderrank algorithm can rank the influence of each author.

But Leaderrank algorithm only considers the influence of authors' collaborators, The influence of certain author also involved with oneself. This paper improve the algorithm, which raise a line to oneself from the certain author. Given as Figure 2.

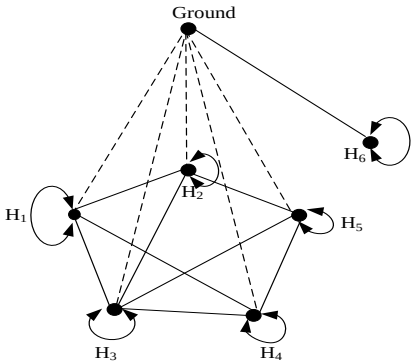

Figure.2. the improved co-authorships diagram

In this case, the influence of certain author is also affected by oneself, the improved formula (1) become formula (2):

$$PR(A) = \frac{1}{C(A)} \times PR(A) + \sum_{i=1}^{n+1} \frac{1}{C(T_i)} \times PR(T_i) \quad (2)$$

Here: $C(A)$ indicates the links of node A.

Assuming that the node T_i connected with node $S_1, S_2, \cdots, S_m$ (a total of m, include A). In formula (1) and (2), the node T_i assigns weights of PR to node A is $\frac{1}{C(T_i)}$, that is node T_i average assigns it's PR to S_i. However, if the original value of node A is different from node S_i, node T_i assigns different weights to A and S_i, It's directly proportional to original value of nodes. The improved Leaderrank algorithm still meets the criteria: (1)More connected node the node A has more important the node A is ,and the $PR(A)$ value bigger; (2) The node A connected with more influential node, that is the value of $PR(T_i)$ is more higher, and T_i assigns more weights to node A. Therefore, the weights of node A assigned to node T_i is defined as:

$$P(T_i) = \frac{PR(A)}{PR(T_i) + \sum_{j=1}^{m} PR(S_j)} \quad (3)$$

The weights of node A assigned to itself is:

$$P(A) = \frac{PR(A)}{PR(A) + \sum_{i=1}^{n} PR(T_i)} \quad (4)$$

Use formula (4) and (5) respectively replace $\frac{1}{C(T_i)}$ and $\frac{1}{C(A)}$ in (2), then we get the improved Leaderrank formula (5) .

$$PR(A) = P(A) \times PR(A) + \sum_{i=1}^{n+1} P(T_i) \times PR(T_i) \quad (5)$$

H-factor is a way of evaluating academic achievement[9], here describe the influence of Erdös1 by referencing H-factor. This paper write the times of cooperation between Erdös1 and Erdös as Rep1, write the times of cooperation between Erdös1 as Rep2, and write the times of cooperation between Erdös1 and Erdös2 as Rep3. The maximal times of three cooperation would be better represents A influence, so this paper defines the improved H-factor as follow:

$$H = \max\{\text{Rep1}, \text{Rep2}, \text{Rep3}\} \quad (6)$$

There are 511 nodes in Erdös1 network, every node can build the $PR(A)$ equations of influence like formula (5), and H-factor as the initial value of $PR(A)$. Use the method of calculating transition probability matrix in Markov chain can get approximate $PR(A)$ solution. This paper apply MATLAB to iterative calculate the 511 equations, when $PR(A)$ value go stability, this value is the influence of author in co-author network. Eventually we get influence ranking of the top 10 and the worst 10 , they are respectively shown in Table 1, Table 2.

Table 1. the top 10 authors

Ranking	1	2	3	4	5
The influence value	87.2	24.2	21.5	17.8	10.2
The author number	10	165	44	479	148
Ranking	6	7	8	9	10
The influence value	10.1	10.0	8.58	5.81	5.14
The author number	378	187	78	440	341

Table 2. the worst 10 authors

Ranking	502	503	504	505	506
The influence value ($\times 10^{-3}$)	8.53	8.46	7.49	7.49	7.49
The author number	279	225	244	477	311
Ranking	507	508	509	510	511
The influence value ($\times 10^{-3}$)	7.49	7.49	7.49	7.49	7.49
The author number	218	146	130	59	14

CHOOSING THE BEST CO-AUTHOR

Through calculating the influence ranking, it's easy to find that: the authors never co-author with anyone except Paul Erdös, which has small influence in co-author network. How to choose the best cooperator is the key to promote their influence.

The best co-author not only relate to own influence, but also his(her) clustering coefficient and node degree[11]. This paper choose the best co-author through considering three aspects as follow, and take the author (No.14) who has minimum influence as an example to study.

(1) Co-authoring with authors who have bigger clustering coefficient

According to references [12], clustering coefficient is defined as:

$$L_i = \frac{2E_i}{k_i(k_i - 1)} \tag{7}$$

Here k_i indicates node i has k_i edges connected with other nodes; E_i indicates the practical edges between k_i node. No.14 co-author with the top 10 authors who have bigger clustering coefficient, his increase of influence value are shown in Table3.

Table 3. the increase of No.14 influence

The co-author number	315	319	187	344	367
Influence ($\times 10^{-3}$)	1.80	2.03	3.97	1.67	1.62
Increase ($\times 10^{-3}$)	1.06	1.28	3.22	0.91	0.88
The co-author number	369	425	453	455	462
Influence ($\times 10^{-3}$)	1.70	1.74	1.58	1.74	1.95
Increase ($\times 10^{-3}$)	0.95	0.99	0.83	0.99	1.20

(2) Co-authoring with authors who have bigger node degree

No.14 co-author with the top 10 authors who have bigger node degree, his increase of influence value are shown in Table 4.

Table 4. the increase of No.14 influence

The co-author number	78	187	341	440	438
Influence ($\times 10^{-3}$)	2.22	3.97	2.12	2.06	1.79
Increase ($\times 10^{-3}$)	1.47	3.22	1.37	1.32	1.04
The co-author number	148	479	44	378	165
Influence ($\times 10^{-3}$)	2.12	2.64	3.62	2.20	3.24
Increase ($\times 10^{-3}$)	1.37	2.89	2.87	1.45	2.50

(3) Co-authoring with influential authors

No.14 co-author with the top 10 influential authors, his increase of influence value are shown in Table 5.

Table 5. the increase of No.14 influence

The co-author number	10	165	44	479	148
Influence ($\times 10^{-3}$)	2.22	3.97	2.12	2.06	1.79
Increase ($\times 10^{-3}$)	3.15	2.50	2.87	1.89	1.37
The co-author number	378	187	78	440	341
Influence ($\times 10^{-3}$)	2.12	2.64	3.62	2.20	3.24
Increase ($\times 10^{-3}$)	1.04	3.22	1.47	3.12	1.47

In conclusion, the minimum influence author (No.14) had better co-author with No.187, then his increase of influence is biggest and it's 3.22×10^{-3}. That is, ASHBACHER, CHARLES D. had better cooperate with ECKLUND, EARL F., JR.

However, No.14 not got the biggest increase after cooperating with No.10 (who is the most influential author in co-author network except Paul Erdös). This is because those co-authors around No.187 have more influential than No.10. Thus it can be seen that perhaps the best co-author is not the most influential author in co-author network. Therefor, authors need to consider influence, clustering coefficient and node degree when they choose the best co-author.

CONCLUSION

The analysis of Paul Erdos co-author network shows that: a certain author's influence not only effected by the times of cooperation with co-authors, but also oneself. So the improved LeaderRank algorithm with H-factor is more suitable to measure the influence of authors than original LeaderRank algorithm.

In order to promote the influence of authors, this paper provide three ways to choose the best cooperator. The practical example indicates that: three ways are reasonable and effective, authors can find the best way to promote his(her) influence through comparison in different circumstance.

REFERENCES

[1] Liu X M, Bollen J, Nelson ML, et al, 2005, "Co-authership networks in the digital library research community", Information Processing and Management, No.41, pp 1462-1480.

[2] Zuo Meiyun, Wen Xiaowei, Hua Xiaoqing, 2012, "Analysis on Co-authorship Network of Scholars in Knowledge Management", Journal of Information Resources Management, No.4, pp 4-15.

[3] Linyuan Lü, Yi-Cheng Zhang, Chi Ho Yeung, et al, 2011, "Leaders in social networks, the delicious case", PloS one, vol.6, No.6, pp 1-9.

[4] LI Liang, ZHU Qinghua, 2008, "An Empirical Study of Coauthorship Analysis Using Social Network Analysis", Information Science, vol.26, No.4, pp 549-555.

[5] ZHANG Fuzeng, YANG Hongyong, LI Ali, 2007, "Scientists Relation Networks and Scientists Influences", Complex Systems and Complexity Science, vol.4, No.2, pp 45-49.

[6] Erdos1, Version 2010, October 20, 2010 https://files.oakland.edu/users/grossman/enp/Erdos1.html

[7] Li Qian, Zhou Tao, Lv Linyuan, 2003, "Identifying Influential Spreaders by Weighted LeaderRank", pp 47-55.

[8] JM Kleinberg, 1999, "Authoritative Sources in a Hyperlinked Environment", Journal of the Association for Computing Machinery, vol.46, No.5, pp 604-632.

[9] Hirsch JE, 2005, "An index to quantify an individual's scientific research output", National Academy of Science of the United States of American, vol.102, No.46, pp 16569-16572.

[10] SHAO Jingjing, FENG B0, LI B0, 2008, "PageRank algorithm ranking of the new technology", Journal of Huazhong Normal University(Natural Sciences), vol.42, No.4, pp 504-508.

[11] Dorogovtsev S N, Mendes J F F, 2002, "Evolution of networks", Advances in Physics, vol.51, No.4, pp 1079-1187.

[12] Réka Albert, Albert-László Barabási, 2002, "Statistical mechanics of complex networks", Reviews of Modern Physics, vol.74, No.1, pp 47-97.