Analysis of calculating authors' breadth centrality in the collaboration network

Guangming Li*, Lu Dai

College of computer, Dongguan University of Technology, Dongguan 523808, Dongguan

**Corresponding author's e-mail: dailudgut@163.com*

Received 25 December 2013, www.cmnt.lv

Abstract

In collaboration network, based on the relations and cooperation between coauthors, we evaluate importance of scholars, in what position scholars are, such as the centrality of scholars. This information reflects scholars 'research cooperation. In this paper, the methods were given up from various perspectives, which indicate that the network contains a lot valuable information about scholars. Mining and identification of this information with in-depth analysis will play a significant role in guiding the formulation of science and technology management and technology policy.

Keywords: collaboration network; complex network community; breadth centrality

1 Introduction

As the development of science and technology, the research gets more and more sophisticated .it is hard for researchers to complete a research or thesis alone. Scientific publications, as main output of scientific research, co-published by several authors are much more general, rather than single-authored publications [1]. Collaboration-network, which consists of the author with the edge as the Collaboration-relationship reflects the Collaboration between experts and scholars in scientific publications [2]. Analysis of this network can reveal a lot of information about Collaboration structure between scholars, for example, who are co-authors of a scholar, whether a scholar in a core of this network, which co authorships are stronger than others, What is the difference between their role in the network.

Firstly, based on the quantitative graph theory to measure and analyze these qualitative changes in the cooperation network topology, such as density, diameter, and relative size of the largest component of the network. he analysis shows that the development of a field would experience structure topology evolution from relatively small disconnected components to large networks with huge connected component. Therefore, the number of edges and nodes in the largest component will experience evolution from relatively few to the many.

Firstly, the methods of calculating authors' breadth centrality were introduced. Considering that the distribution of cooperative relationship can be used to demonstrate the authors' cooperation distribution in different subgroups, thus we can calculate the breadth of the authors' cooperative relationship through the distribution of the cooperative relationship, i.e. breadth centrality [3]. Those authors of even cooperative relationship distributions in multiple subgroups have high breadth centrality; while those of uneven cooperative relationship distributions in multiple subgroups have relatively lower breadth centrality; one extreme case is that the authors' distribution of cooperative relationship in different subgroups are absolutely even, in which case, the breadth centrality reaches up to the maximum. Another extreme case is that authors have only cooperated with the authors subjecting to their own subgroups; that's to say, the cooperative relationship is confined to their own subgroups; in this case, the breadth centrality reaches the minimum, i.e. 0.

2 The calculation methods of author breadth centrality

The idea of calculating the breadth centrality was worked out under the inspiration of Tutzauer's entropy centrality and Borner's studies [4, 5]. However, it must be emphasized that there is significant differences between the breadth centrality method and the above two. The studies of Tutzauer and Borner proposed that the distribution is related to the individual channel or relationship, i.e. their studies mainly focused on the distribution of relationship with other individuals. The breadth centrality aims to analyze if the authors have been cooperated with others in the multiple subgroups, as well as others within and beyond their own subgroups.

Although there are significant differences between the breadth centrality method and Tutzauer's entropy centrality and Borner's studies, there is a very important point in common between them, i.e. the application of the principle of entropy calculation. According to the characteristics of the breadth centrality stated above, the specific calculating methods of which can be obtained. It is based on the application of the principle of entropy calculation as well. Therefore, the specific calculating methods are stated as follows:

Assuming that there are n authors and c subgroups in one coauthor network, thus the sum of cooperative relationship strength of one author i and the author from the k^{th} subgroup is

$$sum_k(i) = \sum_j rs_{ij} \ . \tag{1}$$

Among the formula, "j" is one of the cooperator of the author i in the k^{th} subgroup;

rs_{ij} is the cooperative relationship strength between author i and author j calculating by the basic Salton method.

Thus, the proportion of the cooperative relationship strength between authors i and the author in the k^{th} subgroup

in the total cooperative relationship strength of author i can be demonstrated as the following formula:

$$p_k(i) = \frac{sum_k(i)}{\sum_{j=1}^{n} rs_{ij}} . \quad (2)$$

Based on the principle of entropy calculation proposed by Shannon, the breadth centrality of author i is:

$$C_{ext}(i) = -\sum_{k=1}^{c} p_k(i) \log p_k(i) . \quad (3)$$

There is one shortcoming in formula 5-5, i.e. it excluded a ubiquitous case. Generally, one author may have been cooperated with authors from one or multiple subgroups, but other cooperators may not belong to any subgroup. In order to solve this problem, we included the authors that do not belong to any subgroup into an individual subgroup-"third party". In order to demonstrate this point in formula (5-5), the formula was altered slightly, thus a more accurate formula for calculating the breadth centrality is set as follows:

$$C_{ext}(i) = -\sum_{k=1}^{c} p_k(i) \log p_k(i) - (1 - \sum_{k=1}^{c} p_k(i)) \log(1 - \sum_{k=1}^{c} p_k(i)) . \quad (4)$$

In this paper, the base of logarithm is set as 2. In addition, what must be particularly stressed is that if the author and the author's cooperator belong to a same subgroup, or do not belong to any subgroup, thus the breadth centrality of the author is 0.

3 The experiment of complex network community detection methods

Authors can be divided into "special experts" and "integrated experts" according to their degree centrality C_d, number of research directions N_s and heterogeneity of research directions H_s. According to the definition of the breadth centrality, authors of high C_{ext} have been cooperated with the authors from multiple subgroups, or the internal and external authors belonging to the same subgroup, thus such authors work as a "bridge" between multiple subgroups, or in the "external exchange interface" position of the subgroups they belong to. They can promote the exchange of information, knowledge, thinking model, methodology, and other abundant resources between different subgroups. This kind of author is called as "bridge experts".

Here is a complementary point: the author of high breadth centrality can work as an "introducer", but the research directions of different subgroups should be taken into consideration. For convenience of explanation, here take constituent element 3 for instance.

FIGURE 1 2014 subnets in experimental data for the division of the sample

Shared by many practical of complex network modeling system of the emergence of a feature, the clustering network nodes tend to through the internal connection of many communities (and not just by chance would like) and some other external connection form groups in the community. Social and information networks are such examples of some types of network. Recently, many studies have shown that this kind of actual complex networks may show the local features, this and the global features of the entire network is has essential difference, so if the network as a does not take into account the community structure of the whole may overlook certain qualities in many modeling system. [6] In This article, we will use the method based on spectral decomposition to find community [0]. For simplicity's sake, let us consider the case, the network is composed of two communities. Set R is the number of community given by.

$$R = \frac{1}{2} \sum_{G_i \neq G_j} A_{ij} . \quad (5)$$

Among them, said only when the nodes i and j is considered to be placed in the sum of different communities. Each node is stored in the relationship between including vector of n elements. If the node I belong to the community the G_1 node, and then $S_i = 1$. Otherwise, $S_i = 1$, I belongs to the G_2. Please note that if we will introduce a. formula 1, can be rewritten as:

$$R = \frac{1}{4} \sum_{ij} (1 - s_i s_j) A_{ij} . \quad (6)$$

All degrees of the network is:

$$k = \sum_{ij} A_{ij} = \sum_{i} k_i = \sum_{i} s_i^2 k_i = \sum_{i} s_i s_j k_i \delta_{ij} . \quad (7)$$

Using A formula a. 2. 4 replacement, we get:

$$R = \frac{1}{4}\sum_{ij}(s_i s_j)(k_i\delta_{ij} - A_{ij}) = \frac{1}{4}\vec{s}^{T} L \vec{s} \cdot \quad (8)$$

Because of the number of edges between R communities, our goal is to minimize the amount. In order to do this, we will first Laplacian matrix L expanded into a linear combination of the feature vector

$$R = \sum_i a_i \vec{v}_i L \sum_j a_j \vec{v}_j = \sum_{ij} a_i a_j \lambda_j \delta_{ij} = \sum_i a_i^2 \lambda_i \ . (9)$$

Here is the corresponding eigenvalues. In order to minimize the R, we need to ai and minimum eigenvalue corresponds to the highest coefficient. Therefore, the objective function is to put on and parallel direction. Unfortunately, if a person according to the optimization strategy, will be a trivial solution, so all the nodes will belong to a single community. Therefore, compared with the eigenvector, we will feature vector projection to the same direction. In order to minimize the amount of edge, between the community is one of the most inspired to $s_i = 1$, whether or not the node I is negative. Also, si should be set to the $s_i = 1$ regardless of the ith node is positive, minimize between products.

In order to illustrate how algorithm works, we examined the expression of social network in the two communities. As shown, the two is consistent, as we had expected, each community is made up by the edge of many in the community and the community of outside edge. Newman mentioned in the previous section, we put forward the method and the methods described here are very similar, and in addition to minimize the number of edges between community, Newman algorithm also maximize the number of community inside edge [7], a section on the modal function of random expected. At the same time, they provide the method not only to detect more than two communities may existence.

4 Implications for technological management

In technological management, the experts and scholars should be evaluated from the distribution pattern of their research directions, including the number, the heterogeneity and the evenness of their research directions. The implications for technological management that obtained by analyzing the above aspects are as flows:

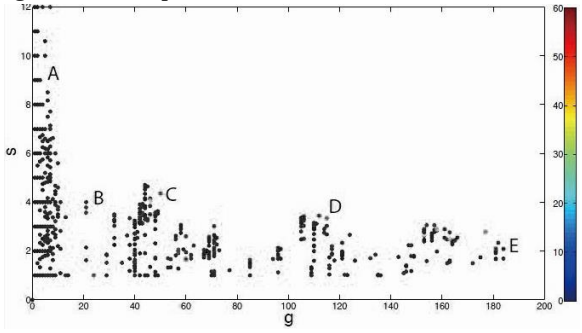


FIGURE 2 Variables gmax (in size) and s (average period of cooperation) between the scattering diagram

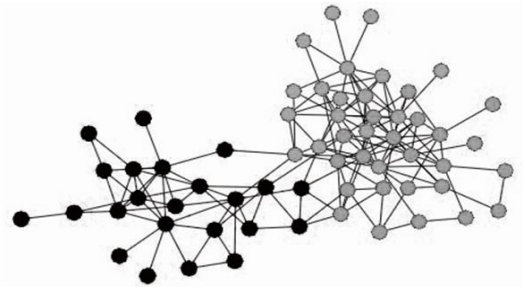


FIGURE 3 Dolphins social network on the basis of the spectral decomposition method for instance.

Firstly, the correlation analysis on authors' degree centrality, the number of authors' research directions and the heterogeneity of authors' research directions indicated that the "special experts" and "integrated experts" should exert different roles in the field. Technological management department often selects suitable experts and scholars to form a Project Review Panel, and this panel may have to review different projects of one field; therefore, this panel should be formed by "special experts" and "integrated experts" [8, 9]. The "integrated experts" act as the leaders of panel to organize and coordinate the reviewing of the whole project, while the "special experts" exert their professional knowledge to have in-depth review on the project. In addition, the research group assuming interdisciplinary and integrated projects should be formed as "special experts" and "integrated experts" as well, which is a necessary factor that must be considered in the management of scientific researches.

Secondly, the correlation analysis on authors' degree centrality, the number of authors' research directions and the heterogeneity of authors' research directions indicated that although some authors have a few research directions, they have high breadth centrality. [10] More importantly, the "bridge" experts sometimes could act as the introducers between different subgroups, and sometimes they belong to the "introduction-type" experts. When two subgroups own a common research direction but have not cooperated in this common direction, the "introduction-type" experts should be encouraged to introduce the authors from different subgroups to have cooperative research on this direction [11].

Thirdly, the authors of highly even distribution of research directions reasonably allocated their different research directions to different cooperators, instead of limiting to a certain part of authors [12]. Therefore, this kind of allocation is not only significant for improving the efficiency of research and widely accepting all kinds of information and knowledge, but also for efficient distribution and planning of research directions. Apparently, this kind of cooperation model is very worthwhile to promote, and in technological management work, experts should be encouraged to use this cooperation model as well [13, 14].

5 Conclusions

As the result of the Analysis, some interesting trends in the of scientific cooperation are obtained, such as that the average size of the group grows exponentially, and the increasing of number of authors obeys the power-law, and by extrapolation it is able to determine the approximate date of a separate affine group. Also, spectral analysis-based approach can be used to divide communities in the network.

These studies not only enriched the theory of collaboration network, also provides useful lessons for science and technology management, and science and technology policy formulation.

References

[1] Lusseau D, Schneider K, Boisseau O J, Haase P, Slooten E, Dawson S M 2009 The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait? *Behavioral Ecology and Sociobiology* **54**

[29] Amancio D R, Nunes M G V, Oliveira Jr, O N, Costa L da F 2011 Using complex networks concepts to assess approaches for citations in scientific papers *Scientometrics* **90**(2)

[30] Amancio D R, Oliveira Jr, O N, Costa L da F 2012 Three-feature model to reproduce the topology of citation networks and the effects from authors visibility on their h-index *Journal of informetrics* **6**

[31] Hsiang L C, Rebecca Y H 2012 Quantifying the degree of research collaboration: A comparative study of collaborative measures *Journal of Informetrics* **6**(1)

[32] Mieghem P 2011 *Graph Spectra for Complex Networks* Cambridge University Press

[33] Perra N, Gonc alves B, Pastor-Satorras R, Vespignani A 2012 Activity driven modeling of time varying networks *Scientific Reports* **2** 469

[34] Costa L da F, Rodrigues F A, Travieso G, Villas Boas P R 2007 Characterization of complex networks: a survey of measurements *Advances in Physics* **56** 167-242

[35] Ding Y 2011 Community detection: Topological vs. topical *Journal of Informetrics* **5**

[36] Girvan M, Newman M E J 2002 *Community structure in social and biological networks* Proc. Natl. Acad. Sci. USA **99**

[37] Persson O 2013 Identifying research themes with weighted direct citation links *Journal of Informetrics* **4**(3)

[38] Roth C, Wu J, Lozano S 2012 Assessing impact and quality from local dynamics of citation networks *Journal of informatrics* **6**(1)

[39] Newman M E J 2009 Coauthorship networks and patterns of scientific collaboration *Proc. Natl. Acad. Sci. USA* **101**

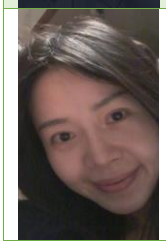
[40] Newman M E J 2010 The structure of scientific collaboration networks *Proc. Natl. Acad. Sci. USA* **98**

[41] Shrum W, Chompalov I, Genuth J 2001 Trust, conflict and performance in scientific collaborations *Social Studies of Science* **31**(5)

Authors

Li Guangming, born in 1968

Current position, grades: vice professor at Dongguan University of Technology. His
University studies: B.S degree form Henan Normal University in 1995, and the M.E. degree form Zhengzhou University of Technology in 2000.
Scientific interests: nonlinear circuit theory, chaotic signal processing, embedded systems design, social network and their applications.



Dai Lu, born in 1988

Current position, grades: working at DongGuan Univercity of Technology.
University studies: Ph.D. degree from Wuhan University, in 2013.
Scientific interests: cloud computing, evolutionary algorithm, social network and text mining.