

# NODE IMPORTANCE MEASURE FOR SCIENTIFIC RESEARCH COLLABORATION FROM HYPERNETWORK PERSPECTIVE

*Quan Xiao*

Original scientific paper

Collaboration has become main stream and trend in interdisciplinary fields. In research collaboration organizations, to evaluate the contributions of researchers to the organization and then to identify core researchers is an important issue to carry out performance appraisal and crisis management of brain drain. Scientific research collaboration network is a basic model to investigate this question, but under the context of increasingly complex collaborative behaviour, it shows its limitations for semantic representations. In this paper, by introducing hypernetwork, a more powerful modelling tool than traditional network, and taking scientific paper co-authorship as object to construct scientific research collaboration hypernetwork (SRCH), we measure the importance of researchers in two aspects, as collaborative relationship structure and collaborative achievement value from a hypernetwork perspective. An additive weighting method with adjustable parameters is utilized to integrate the evaluation indicators of the two aspects, and then the synthetical importance evaluation of researchers is obtained. Analysis of data instance verifies that our node importance measure for scientific research collaboration from hypernetwork perspective is reasonable and effective.

**Keywords:** adjustable parameter; hyper network; node importance; scientific research collaboration network

## Mjera važnosti čvora u suradnji pri znanstvenom istraživanju sa stajališta hiper mreže

Izvorni znanstveni članak

U interdisciplinarnim područjima sve se više pažnje posvećuje međusobnoj suradnji. U organizacijama koje se bave istraživanjem suradnje, od bitne je važnosti evaluirati doprinos istraživača svojoj organizaciji te identificirati glavne istraživače. Mreža suradnje u znanstvenom istraživanju predstavlja osnovni model, ali u kontekstu sve složenijeg suradničkog ponašanja javljaju se problemi oko njegovog semantičkog predstavljanja. U ovom radu, uvođenjem hiper mreže, moćnijeg alata za modeliranje od tradicionalne mreže i uzimajući koautorstvo u znanstvenom radu kao cilj za stvaranje hiper mreže suradnje u znanstvenom istraživanju (scientific research collaboration hypernetwork - SRCH), mjerimo važnost istraživača u dva smjera, kao strukturu suradničkog odnosa i vrijednost postizanja suradnje sa stajališta hiper mreže. Upotrebljena je dodatna metoda mjerenja s prilagodljivim parametrima u svrhu integracije indikatora evaluacije dvaju aspekata te je dobivena procjena sintetičke važnosti istraživača. Analiza rabljenih podataka potvrdila je da je naša mjera važnosti čvora u znanstveno istraživačkoj suradnji sa stajališta hiper mreže razumna i učinkovita.

**Ključne riječi:** hiper mreža; mreža istraživanja suradnje u znanstvenim istraživanjima; prilagodljivi parametar; važnost čvora

## 1 Introduction

Under the support of information technology, the rapid development in interdisciplinary fields makes deep collaboration necessary and possible, while collaboration has been becoming the mainstream ways for scientific researches. In scientific research organizations, core researchers are those who own key competences and are indispensable for the organization to implement innovation strategy or to build scientific research alliance. To evaluate and identify these talents is a crucial issue for organizations to implement performance appraisal and management of knowledge employee, and it is also a key problem to be solved for the crisis of brain drain.

Recently, with growing maturity of theories, methodologies and technologies of network analysis, to investigate big data by network has become a hotspot in many disciplines [2]. The collaboration relations between scientific research organizations or researchers undisputedly form a complex network, which is composed of knowledge, researchers, achievements, carriers and organizations themselves. Consequently, in the complex network of scientific research collaboration, we can take advantage of network analysis methods to describe the relations of paper citations, researchers' co-authorships, and the influence of each other in researcher network [5].

In the context of complex collaboration behaviour, in this paper we try to introduce hypernetwork, a tool with more powerful modelling capability than ordinary network, and evaluate the importance of researchers in

scientific research organizations in two aspects of collaborative relationship structure and collaborative achievement value from a hypernetwork perspective. Concretely, taking scientific research paper co-authorship as object to construct scientific research collaboration hypernetwork, we measure the node importance by weight information of nodes and hyperedges, i.e., quantitatively analysing and evaluating the academic standings, academic contributions, scientific research collaborative betweenness, and output competences of all authors integratedly, which finally provide evidences for the identification and evaluation of core talents in research organizations.

## 2 Related works and review

### 2.1 Scientific research collaboration network

Scientific research collaboration network is a kind of social network particularly for describing the collaborative relationships of research participants (researchers or research organizations). In scientific research collaboration network, a research participant is usually represented as a node, if there exists collaborative relation between two different participants, then a line is drawn connecting the corresponding two nodes. Thus, a network with complex structure and dynamic evolution characteristics is formed.

Reviewing literature of scientific research collaboration network on different aggregation level, there are researches on collaborations in different fields

on personal level [14], and researches on organizational level, which investigate collaboration inside an organization [15] or cross-organizations [9]. In addition, according to research focus, overall network and individual network researches are included. Overall network mainly concerns the whole topological structure of knowledge or social network formed by researchers and the impact of its evolution on knowledge forming, spreading, and recombination both inside and cross organizations. For example, Yan [13] used data from 18 core source library and information science (LIS) journals in China covering 6 years to identify the collaboration pattern and network structure of the co-authorship network of LIS in China. Individual network based primarily on the fact that "people tend to discover information, attain knowledge and methods of solving problems relying on their own social network", intensively investigates the centrality and tendency of individuals in network. Thus Singh [9] examined how collaboration across national, organizational or institutional boundaries contributed to knowledge creation, and discovered that external collaboration did significantly improve future publication productivity of the collaborating scientists.

Currently, there appear more and more theories and techniques applying on the modelling and analysis of scientific research collaboration network, e.g. complex network [1], cluster analysis [12] and Web mining etc.

## 2.2 Node importance measure in network

In a network, to evaluate the importance of nodes and identify the key ones has deserved the research in fields of network analysis and system sciences [17]. In recent years, the blossom of empirical study in network, especially the discovery of "small world" and "scale-free", makes it reasonable to evaluate the importance of network nodes on the basis of their dissimilarity. The method of measuring node importance in network arisen at present within different application background can be classified into two types as follows.

### (1) Social network analysis (SNA) methods

Major methods for measuring node importance based on SNA are under the assumption that the importance of a node is equivalent to its notability when connecting with other nodes (Knoke and Burt, 1983). These methods generally study the metrics while maintaining the integrity of network, i.e., do not break the connectivity. The basic idea is to find out some useful attributes (such as quantity of information contained in degree or shortest path) to highlight diversity between nodes, in other words, to adequately display the positional traits of nodes in a network, and to "magnify" the significance to define the importance. Proposed metrics based on SNA include two categories as "centrality" and "prestige", while indicators include degree, closeness, betweenness, eigenvector, cumulated nomination etc. Many representative works have been done with SNA based methods by researchers (Bonacich, 1987; Altmann, 1993; Poulin, Boily, and Masse, 2000).

### (2) Node deleting methods

It is another thinking to evaluate the importance of a node in network by measuring the destructiveness to the network performance when removing it. That is

"destructiveness equals importance". The greater the extent of severity when we remove a node is, the more important it is. Because the maintenance of network connectivity or system function depends on its existence. Thus Xi [16] took the inverse of distance between node pair as weight value, and then calculated the weighted sum of all disconnected node pairs to measure the destructiveness to the network connectivity. This work made a quantized contribution to the "removing based" thinking. In general, it is necessary for node deleting methods to settle the problems of topological changings, even the segmentation of the whole network caused by the nodes removed.

From above we see that single measuring indicator will depend seriously on a single side feature of network, thus ignore the significance of other indicators. Therefore in practical applications, multiple methods and indicators are adopted to measure the importance of nodes in one network corporately.

### (3) Analysis of node importance measure in SRCH

As a typical kind of network, SRCH and its node importance measuring have attracted much attention of researchers. Newman [6] studied a variety of nonlocal statistics for these networks such as typical distances between scientists through the network, and measures of centrality such as closeness and betweenness, and found that typical distances between pairs of authors through the networks were small-the networks form a "small world" in the sense discussed by Milgram. Tang [11] developed a data mining tool for scientific literature called ArnetMiner, in which a function of discovering important authors was provided. ArnetMiner mainly took the times of collaboration between authors into account.

A lot of achievements have been emerging about SRCH and its node importance measuring, however, there are still some problems that need to be resolved.

- The representation of network is based on the classical graph theory, which is incompetent to express the characteristic of multi element, multi-layer or multi-granularity. Moreover, most of researches consider connected network, many measuring methods cannot support unconnected network with effect.
- Each single indicator focuses on one particular feature of network, e.g. the position of node in network structure, the influence and control capability of node to information spreading, or the contribution of node in substructure. Specifically, degree only emphasizes the number of edges connected with adjacent nodes, which ignores the indirect effects from them. Betweenness is used to describe the informational control capability of a node to the connection between other nodes, but it is unable to mediate its partial contribution. Removing based methods are difficult for measuring the node importance when the influence of removing a node to the network performance is too high or too low.
- It is requisite to traverse the whole network for the node importance measuring based on overall perspective. For instance, the calculation of betweenness is feasible in a small network, but many networks in reality are of substantial scales and complex structures, which makes it impossible to measure node importance from overall perspective.

### 2.3 Hypergraph and hyper network

As analysed above, traditional network based on classical graph theory has been increasingly manifesting its modelling limitations with the increase in network complexity. The emerging of hypergraph and hypernetwork/supernetwork, provides a new research view for different kinds of networks and their relationships. The concept of hypernetwork was first

proposed by Sheffi [8] in 1985, which is defined as "a network beyond existing network" and "network in network" (Sheffi [8], Nagurney [4]). If a network is represented by hypergraph, it can be deemed as hypernetwork. Hypergraph and hypernetwork take the advantage of representing the characteristics of nesting, multi-layer, multi-level, and multi-attribute that are beyond the capability of ordinary graph and network.

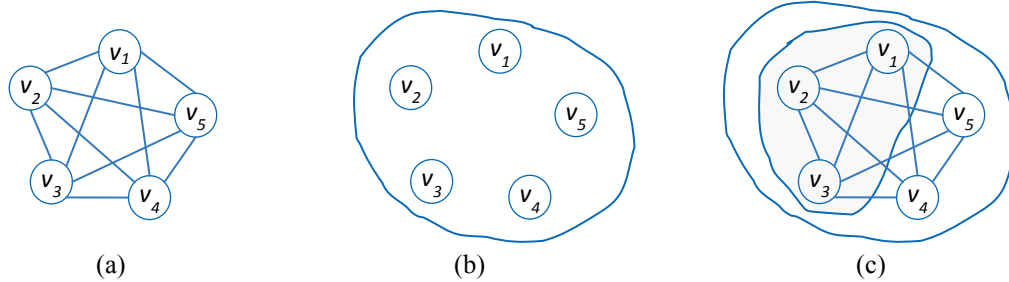


Figure 1 Hypernetwork representation of scientific paper co-authorships

Taking SRCH to be researched as an example, an author is described as a node, and a paper written by two or more than two authors can be represented as one hyperedge connecting more than two nodes. Although a hyperedge with many nodes can be transformed to ordinary network that all nodes connected with each other, but in this case it is difficult to distinguish whether it is co-author in pairs or multi-author collaboration. The ordinary network shows its limitation for semantic representations.

As shown in Fig. 1, the case of five authors writing one paper collaboratively can be represented as hypernetwork as Fig. 1b. Fig. 1a shows the case that each two of the five authors collaborated, in which way 10 papers are written totally. In Fig. 1c, each two of the five authors collaborated, and the five authors wrote one paper collaboratively, and what is more, authors 1, 2, and 3 wrote another paper collaboratively. Totally, there are 12 papers in Fig. 1c. If we analyse by degree of nodes, in Fig. 1a and Fig. 1b, the degree of all nodes is 4, but its implication is completely different. Therefore, metric indicators in classical graph cannot be transplanted directly to hypernetwork, which takes more advantage of concise and precise in representation than classical graph as a modelling method.

Actually in hypernetwork representation, hyperedge does not have to describe co-authorship relations, it is also able to depict relations of emotion of trust relations by setting different attributes, through which "homogeny nodes and heterogeny hyperedges" can be achieved. As a research tool, hypernetwork model can be utilized to express the interactions and influences between networks. So far, researches on hypernetwork mainly concentrated on the construction of supply-chain hypernetwork, finance hypernetwork, etc.

### 3 Measuring method from hyper network perspective

#### 3.1 Symbol definition

To facilitate the description of our method, we give the definitions that appeared in the measuring model first as bellow:

#### (1) Definition of hypernetwork

We refer to Berge's basic definition of hypergraph (Berge, 1973) to define hypernetwork as follows:

Definition 1: A hypernetwork  $HN = (V, E)$  on a set  $V$  is a family  $e_j$  of non- empty subsets of  $V$  called hyperedges, where  $V = \{v_1, v_2, \dots, v_n\}$  a finite set of nodes is,  $E = \{e_1, e_2, \dots, e_m\}$  is a family of subset of  $V$ , s. t.  $e_j \neq \Phi (j = 1, 2, \dots, m)$  and  $\bigcup_j e_j = V$ . In hypernetwork

$HN$ , elements  $v_1, v_2, \dots, v_n$  of set  $V$  are called node, while  $e_1, e_2, \dots, e_m$  are called hyperedge.  $Card(V)$  and  $Card(E)$  are denoted as the number and nodes and hyperedges of  $HN$  respectively.

The hypernetwork representation of the three graphs in Figs. 1a ÷ 1c is:

$$HN_1 = \{\{v_1, v_2\}, \{v_1, v_3\}, \{v_1, v_4\}, \dots, \{v_4, v_5\}\},$$

$$HN_2 = \{\{v_1, v_2, \dots, v_5\}\},$$

$$HN_3 = \{\{v_1, v_2, \dots, v_n\}, \{v_1, v_2, v_3\}, \{v_1, v_2\}, \{v_1, v_3\}, \dots, \{v_4, v_5\}\}.$$

Besides the form of set, hypernetwork has some equivalent forms as closed curves, matrices and bipartite graphs. Among them, matrix form has the features of conciseness, systematicness and adaptability in calculating, so hypernetwork is often handled as matrix in scientific or engineering calculation tasks.

#### (2) Matrices of hypernetwork

##### a) Incidence Matrix (IM)

Definition 2: **IM** of hypernetwork  $HN$  is a  $m \times n$  matrix, where the  $m$  rows of **IM** correspond to the  $m$  hyperedges  $e_1, e_2, \dots, e_m$  of **IM**, and the  $n$  columns correspond to the  $n$  nodes  $v_1, v_2, \dots, v_n$ . If  $v_i \notin e_j$ , then element  $im_{ji} = 0$ , else if  $v_i \in e_j$ , then  $im_{ji} = 1$ .

##### b) Adjacency Matrix (AM)

Definition 3: In hypernetwork  $HN = (V, E)$ ,  $\forall$  vertices pair  $\langle v_i, v_j \rangle \in V$ , if  $\exists e_k \in E$  s. t.  $\{v_i, v_j\} \subseteq e_k$ ,

then  $\langle v_i, v_j \rangle$  are adjacent in  $HN$ ; for vertices pair  $v_i, v_j \in V$  and a subset  $E'$  of  $E$ , s. t.  $\forall e_k \in E'$ , have  $\{v_i, v_j\} \subseteq e_k$ , and  $\forall e_l \in E - E'$ , have  $\{v_i, v_j\} \not\subseteq e_l$ , denote  $Card(E') = \delta$ , then the adjacency degree of  $\langle v_i, v_j \rangle$  in  $HN$  is  $\delta$  (denote  $AD(v_i^i, v_i^j) = \delta$ ). Elements of Adjacency Matrix  $AM_{n \times n}$  of  $HN$  is  $am_{ij} = AD(v_i, v_j)$ .

**c) Distance Matrix (DM)**

Definition 4:  $DM_{n \times n}$  is the Distance Matrix of hypernetwork  $HN$ , where element  $dm_{ij}$  in  $DM_{n \times n}$  is the shortest hyperpath from node  $v_i$  to node  $v_j$ .

**3.2 Basic ideas for measuring method**

Although numerous indicators and methods are proposed according to different standards to measure which nodes are more important than the others, in SRCH, what should be deemed as the criterion is the scientific research capability, which is usually measured by the quantity and quality of outputs as co-authored papers, books or patents for inventions. Wherein, co-authorship relations of scientific research papers are easy to acquire and of strong explanations (Melin [3]), so it is a powerful tool to adopt social network analysis methods to investigate the phenomenon of scientific research collaborations by co-authorship relations (Newman [5]).

In many researches of scientific research collaboration network, node importance depends on the degree of node. In this way, author with most collaborators is obtained, but it is inadequate to judge the comprehensive influence of him. In the following some typical viewpoints are given. First, node importance in scientific research collaboration network should integrate weight information of nodes and edges into account, where weight of nodes reflects contribution of authors and weight of edges reflects value of papers. Second, core author in a discipline should meet the two conditions: he has published plentiful papers and, he has collaborated with others broadly.

Therefore, when studying the importance of researchers to organization, we should consider the extensive degree of collaboration with others, i.e. collaborative relationship structure and collaborative achievement value.

**3.3 Measuring methodology**

**(1) Measure in collaborative relationship structure**

In research of (Xiao [17]), we discovered that measuring methods based on classical graph are inapplicable to hypernetwork. Besides, centrality will lose efficacy in unconnected network, so it is necessary to reconsider the problem of measuring collaborative relations under hypernetwork model. Xi and Tang [16] proposed the concept of multiplex multi-kernel network, and took over the idea of "destructiveness is equivalent to importance" in the measuring of node(s) importance. The formal description of multi-kernel network is

hypernetwork. So we can introduce the destructiveness idea to measuring node importance in hypernetwork.

For an undirected connected node-weighted network, if a node  $v_i$  is deleted, the connectivity of network may be broken in two aspects: firstly, nodes originally connected with  $v_i$  will be disconnected, and their weight cannot spread to other nodes, which will result in Direct Loss ( $DL$ ); secondly, paths between some of the remainder nodes may be interrupted because of the loss of function as bridge of deleted node  $v_i$ , so these nodes cannot exchange or share weights with each other, resulting in Indirect Loss ( $IL$ ).

Therefore, the importance of node  $v_i$  in hypernetwork  $HN$  is the Total Loss ( $TL$ ) to  $HN$  of deleting it, and calculated as:

$$TL(v_i) = DL(v_i) + IL(v_i) = \sum_{j \neq i} \gamma(v_i, v_j) \times \beta(v_i, v_j) + \sum_{s, t \neq i} \gamma(v_s, v_t) \times \beta(v_s, v_t),$$

where  $\gamma(v_i, v_j)$  is the decay coefficient of collaborative behaviour of  $v_i$  and  $v_j$ , which is defined to be inversely proportional to their distance, because in hypernetwork the effect of farther nodes will be smaller;  $\beta(v_i, v_j)$  is the information quantity between  $v_i$  and  $v_j$ . In this paper, we take as the times of collaboration (i.e. adjacency degree) between them, and  $\beta(v_i, v_j) = \delta$ .

By Adjacency Matrix and Distance Matrix representation, the collaborative relationship structure importance of  $v_i$  can be calculated as:

$$TL(v_i) = \sum_{j \neq i} \frac{am_{ij}}{dm_{ij}} + \sum_{s, t \neq i} \frac{am_{st}}{dm_{st}}$$

**(2) Measure in collaborative achievement value**

Dimension of collaborative achievement value reflects the quantity and quality of outputs by a researcher. On the basis of measuring each paper's academic value and investigating the contribution of each collaborator in it, we accumulate the contributions in all papers group by each author, thus the collaborative achievement value of everyone is obtained.

Generally, the value of a scientific research paper can be evaluated by factors as index source, time cited, impact factor of journal and fund project level, etc. Considering these factors synthetically, we define the value of a scientific paper  $p$  as:

$$v_p = [(1 + \alpha_{r_1}(p)) \times (1 + \beta_{r_2}(p)) + v_{if}(p)] \times \times [1 + \frac{\gamma(p)}{y(p)} \times 0, 1] / 10$$

where,  $r_1 = 1, 2$  and  $r_2 = 1, 2, 3, 4$ .  $\alpha$  represents the index source,  $\alpha_1$  is SCI index,  $\alpha_2$  is EI index, and  $\alpha_1 > \alpha_2$ ,  $\alpha_1 + \alpha_2 = 1$ ;  $\beta$  represents fund project level,  $\beta_1$  is national project,  $\beta_2$  is provincial and ministerial project,  $\beta_3$  is municipal project,  $\beta_4$  is horizontal



project, having  $\beta_1 > \beta_2 > \beta_3 > \beta_4$  and  $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$ ;  $v_{if}$  is the journal impact factor;  $\gamma$  is the number of cited times of the paper;  $y$  is the fixed number of years the paper published. As the cited times will increase with the fixed number of years the paper published, here is the average cited times per year. The formula above demonstrates the impact of index source, time cited, impact factor of journal and fund project level on collaborative achievement value, and the expansion effect of time cited.

In a scientific paper, authors' contribution is different. To describe the contribution proportion in papers with different author numbers, we define an author contribution proportion matrix denoted by  $A$ .  $A$  is a lower triangular matrix, where each row represents situations of different author numbers. Element  $a_{ij}$  is the contribution of author  $j$  in total  $i$  authors. For each situation  $i$ , s. t.  $\sum_j a_{ij} = 1$ . If paper  $p$  totally has  $n_p$  authors, then the

contribution of author  $j$  in the paper is  $C_{pj} = a_{n_p j} \times v_p$ . If author  $j$  published  $s$  papers altogether, then his collaborative achievement value measures as:

$$Val(j) = \sum_{p=1}^s a_{n_p j} \times [(1 + \alpha_{r_1}(p)) \times (1 + \beta_{r_2}(p)) + v_{if}(p)] \times [1 + \frac{\gamma(p)}{y(p)} \times 0,1] / 10.$$

**(3) Comprehensive evaluation for multiple indicators**

Through the analysis of merits and demerits of existing evaluating methods, we know that single indicator is inadequate to reflect the influence of a node to the whole network, so multiple indicators are adopted in our research. However, another problem will be faced, that is how to integrate these indicators to make effective evaluation. Multiplication operator is able to integrate two indicators, but it is easy to amplify the influence of single one and cause the evaluation results unreasonable. So we adapt addition operator and "additive weighting" method to get the synthetical importance evaluation, which is able to overcome the drawbacks of single indicator and

multiplication operator in kinds of applications. Furthermore, a measuring method of node importance in SRCH with adjustable parameters is proposed. Denote  $C(v_i)$  is the importance of node  $v_i$  in SRCH, which is calculated as:

$$C(v_i) = \alpha \cdot TL(v_i) + \beta \cdot Val(v_i),$$

where,  $\alpha$  and  $\beta$  are adjustable parameters;  $TL(v_i)$  is the importance of collaborative relationship structure of node  $v_i$  and  $Val(v_i)$  is the importance of collaborative achievement value of  $v_i$ ;  $\alpha + \beta = 1$ . The greatest advantage of this method is that the parameters in the formula can be adjusted whenever necessary according to the actual network application background.

To facilitate the comparison between different networks and eliminate the influence of network scale to the indicators,  $C(v_i)$  needs to be normalized, i.e. making the values distributed in interval  $[0, 1]$  as:

$$C'(v_i) = \frac{C(v_i)}{\max_{j \in N} C(v_j)}.$$

**4 Instance and analysis**

**4.1 Data acquisitions**

Using the measuring method above, we can construct SRCH of specified organization, subject area, or academic journal (group), then decide on the importance of elements in the data set. Following, we take our working unit, Jiangxi University of Finance and Economics (JXUFE) as an instance, to illustrate the application process, results and analysis of the proposed measuring method. We write a web spider program in Java, crawling the paper information of the last five complete years (2009÷2013), in which the affiliation of author is JXUFE, from scientific research databases, including Wanfang Data, Engineering Village, and Science Direct. Altogether we acquired information of 10 211 papers written by 5261 authors.

**Table 1** Sub-networks distribution

NoA	1	2	3	4	5	6	7	8	9	10	11	12
NoSn	2177	327	112	55	32	34	17	8	5	8	3	2
NoA	13	14	15	16	17	18	19	27	33	41	415	421
NoSn	4	4	2	1	1	1	1	1	1	1	1	1

NoA: Number of Authors; NoSn: Number of Sub-networks

**Table 2** Degree distribution of network nodes

D	0	1	2	3	4	5	6	7	8	9	10	11	12	13
NoN	2177	1327	716	364	246	181	89	49	25	21	13	9	8	5
D	14	15	16	17	18	19	20	21	22	23	24	29	33	
NoN	8	3	2	5	2	2	1	1	2	2	1	1	1	

D: Degree; NoN: Number of Nodes

**4.2 Network characteristic**

Through analysis we discover that the SRCH we investigated is non-connected, which consists of independent sub-networks (See Tab. 1).

In Tab. 1, 2177 among 5261 authors never collaborated with others within the recent five years, and 3167 authors are in relatively small research circles ( $NoSn < 4$ ), which may be caused by their research subjects. In the meantime two large networks include 415

and 421 authors respectively, they play great roles in promoting new knowledge and thinking. Fig. 2 is the Log-Log plot of  $NoA$  and  $NoSn$ , which demonstrate power-law traits of the SRCH.

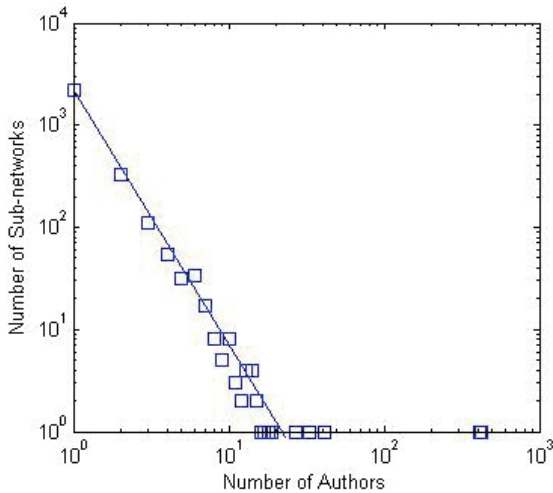


Figure 2 Log-Log plot of  $NoA$  and  $NoSn$

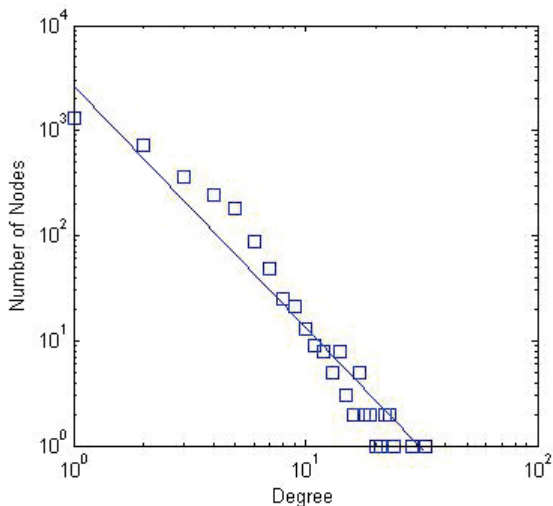


Figure 3 Log-Log plot of  $D$  and  $NoN$

In SRCH of JXUFE, we discover that many nodes have very small degrees (4220 authors have degrees less than 3), they rarely collaborate with others. Only a small number of authors have relatively large degree (53 authors have degrees more than 10). Detailed information is shown in Tab. 2.

According to Tab. 2, the Log-Log plot of  $D$  and  $NoN$  is drawn as Fig. 3, which shows that the distribution of Degree presents power-law decay mode and obvious scale-free characteristic.

Thus, the non-connected phenomenon of collaborative relationships is self-organized, and SRCH is a specific scale-free network.

### 4.3 Measuring results and analysis

#### (1) Measuring results

Tab. 3 below is the measuring result of top 20 important authors at JXUFE measuring through our method.

#### (2) Analysis of results

Through analysis of the collected data, in the last five years on the average level, each author published 2,53 papers while each paper has 1,51 authors, and the average degree of the hypernetwork is 1,62. Taking top 20 important authors into consideration, their degree in average is 18,13, which means each of first-class researchers in this university has a collaboration circle at a scale of about 20; top 20 authors published 25,38 papers in the past five years, which means that if one can publish 5 papers of high quality per year, he will be among the best of staffs. In their papers, there are 2,65 authors in average, which is better than average level within the university.

Moreover, we make some discoveries from data analysis as the following.

- A node with large degree does not always have more synthetical importance or represents an author with superior output capability. Likewise, author with more papers does not mean a more important author. Overall, trends of  $PC$ ,  $D$  and  $TL$  are almost consistent, while deviation also exists to some extent, especially in  $PC$  and  $Val$ , which reflects the discrimination of different level papers on their academic value. Therefore, it is impossible to judge the node importance by only considering single indicator as  $D$  or  $PC$ . However, if bringing  $D$  along with  $PC$  into the evaluation index system, the coupling will be excessive because of the measurement repeatability.
- In the top 20 important authors,  $wsp$  ranked 2<sup>nd</sup> is an extreme example, whose node degree is only 6, which means that he collaborated with only two other authors in the past five years, ranked as No. 384. It is indicated that his collaborative behaviour is non-significant. Among all his papers, 82,31 % are authored by only himself. However, he has published 67 papers in the past five years, ranking first of all, and the  $Val$  of his papers is also the top of all authors. After further investigation, we find that he is a professor in the university and has won the title of "Top 10 Outstanding Researchers" of JXUFE for several years in a row, and has a certain influence in the academic circle of JXUFE. So although the node degree of him has a low value, his ranking as 2<sup>nd</sup> is reasonable. We did an experiment to integrate indicators by multiplication operator, then his synthetical importance was 0,1277, only ranking as 68<sup>th</sup>, which obviously does not conform to the actual situation.
- The  $Val$  of author  $wsp$  is the highest 1, yet that of author  $tcq$  ranking 2<sup>nd</sup> is only 0,3576, a big deviation existing. Through analysis,  $wsp$  has published 18 papers more than  $tcq$ , and their papers' classes are quite. But overwhelming majority of papers of  $wsp$  are authored by himself, while 96,55 % of  $tcq$ 's papers are collaborated with others. Thus the actual contribution rate in each paper of  $wsp$  is higher than that of  $tcq$ , so their gap of 0,6424 in  $Val$  is reasonable. This means the author  $wsp$  has a significant advantage in personal researchability. If we use normalization method that uniformly distributes the ranking into interval  $[0, 1]$ , then the disparity and gap of authors cannot be expressed. This shows that our

measuring method is of preferable explanatory and fault-tolerant capability.

- d) From a macro point of view, there are 1,51 authors in each paper on average, i.e. most papers are accomplished by one or two authors. The average node degree is 1,62, which shows that each author collaborates with 1÷2 partners on average. Among all papers, 70,8 % have only one author, 20,8 % have two authors, only 8,4 % have three or more than three authors, which indicates that the collaborative level is low, and the scientific research management

department should formulate policies to encourage collaborations.

Through contrastive experiments of different node importance measuring methods, the method with adjustable parameters for SRCH proposed in this work is capable of highlighting the dissimilarity between nodes meticulously, and is able to indicate the influences of nodes to the whole network objectively. By adjusting the parameters, this method can be more practical in a wider range, and it can make more guidance value in different network application backgrounds.

**Table 3** Statistical result of top 20 important authors at JXUFE

ID	D	DR	PC	PCR	TL	TLR	Val	ValR	SI	TR
wcx	35	1	44	3	1,0000	1	0,2288	13	0,6144	1
wsp	6	384	67	1	0,1277	934	1,0000	1	0,5638	2
fzj	29	2	28	11	0,9856	2	0,1187	70	0,5522	3
xsh	25	4	39	4	0,8491	4	0,2473	9	0,5482	4
tcq	18	11	53	2	0,7091	11	0,3576	4	0,5333	5
yjh	23	6	26	13	0,8200	6	0,1751	27	0,4975	6
wsq	16	15	14	74	0,8801	3	0,0767	186	0,4784	7
wzy	26	3	16	52	0,8498	5	0,0870	145	0,4684	8
zzh	20	8	30	9	0,6918	16	0,2350	11	0,4634	9
hmh	20	8	21	18	0,7687	7	0,1560	43	0,4623	10
kfb	7	121	17	39	0,5946	43	0,3213	5	0,4579	11
yb	9	61	17	39	0,7124	10	0,1938	20	0,4531	12
ljl	23	6	21	18	0,7673	8	0,1181	72	0,4427	13
why	12	29	18	28	0,7063	12	0,1612	37	0,4337	14
hhp	25	4	11	135	0,7560	9	0,0871	144	0,4215	15
lzj	16	15	18	28	0,7059	13	0,1107	90	0,4083	16
wq	15	18	15	64	0,6897	17	0,1172	79	0,4035	17
tfj	10	46	15	64	0,6088	32	0,1951	19	0,4020	18
leg	20	8	29	10	0,6015	38	0,1789	24	0,3902	19
zr	9	61	12	111	0,5921	45	0,1589	41	0,3755	20

D: Degree; DR: Degree Rank; PC: Paper Count; PCR: Paper Count Rank; TL: Total Loss; TLR: Total Loss Rank; Val: Value; ValR: Value Rank; SI: Synthetical Importance; SIR: Synthetical Importance Rank.

### 5 Conclusions

To identify key nodes in scientific research collaboration network is an important work for scientific research organizations. In view of the limitations of traditional modelling method for scientific research collaboration network, we present a modelling method of hypernetwork, and on the basis of which we investigate the problem of node importance measuring in SRCH in view of hypernetwork. This measuring method with adjustable parameters has the following advantages.

- (1) It is applicable for the evaluation of collaboration within the same organization, or authors of different organizations in the same subject area.
- (2) The parameters in the additive weighting method can be adjusted according to different network application scopes, making it flexible and practical. For example, in early scientific research developing and accumulating period, quantity and quality of outputs should be much accounted for. After some achievements have been made, the structural relations may be concerned.
- (3) We can assign different values of parameters for different nodes in a SRCH, to express the importance of nodes in different combinational value of parameters.

- (4) This method takes some typical indicators into consideration synthetically, avoiding the one-sidedness of measuring based on single indicator.

Although we believe our method is correct and effective, there are still some limitations to be studied and improved in future works such as:

- (1) We take scientific papers as carriers of scientific research collaborative behaviour. In fact, scientific research collaboration is included but not limited to scientific papers. Project cooperation and seminars are also common forms. Additionally, scientific research databases besides Wanfang Data, Engineering Village, and Science Direct can be included as the data source of research.
- (2) Different values of parameters would have different influences on the measuring result, then how to balance the relationships between parameters in accordance with the kind of network, to get maximum effectiveness of the method, is to be resolved.
- (3) In principle and methodology, what we consider is undirected connected network. Whether this measuring method can be generalized to weighted network, directed network or unconnected network needs to be investigated in the future.

## Acknowledgements

This work has been supported by National Natural Science Foundation of China (NSFC, Grant No. 71361011), Science and Technology Project of Education Department of Jiangxi Province (Grant No. GJJ11734) and JXUFE Research Project. The author would like to thank the reviewers for their helpful comments and suggestions for improving the manuscript.

## 6 References

- [1] Albert, R.; Barabasi, A. L. Statistical mechanics of complex networks. // *Rev. Mod. Phys.* 74, (2002), pp. 47-97. DOI: 10.1103/RevModPhys.74.47
- [2] Demirkan, H.; Delen, D. Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. // *Decision Support Systems*. 55, 1(2013), pp. 412-421. DOI: 10.1016/j.dss.2012.05.048
- [3] Melin, G.; Pearson, O. Studying research collaboration using co-authorship. // *Scinetometrics*. 36, 3(1996), pp. 363-377. DOI: 10.1007/BF02129600
- [4] Nagurney, A.; Wakolbinger, T.; Zhao, L. The evolution and emergence of integrated social and financial networks with electronic transactions: a dynamic supernetwork theory for the modeling, analysis, and computation of financial flows and relationship levels. // *Computational economics*. 27, (2006), pp. 353-393. DOI: 10.1007/s10614-006-9031-9
- [5] Newman, M. Scientific collaboration. I. Network construction and fundamental results. // *Physical Review E*. 64, 1(2001), pp. 016131. DOI: 10.1103/PhysRevE.64.016131
- [6] Newman, M. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. // *Physical Review E*. 64, 1(2001), pp. 016132. DOI: 10.1103/PhysRevE.64.016132
- [7] Newman, M. The structure of scientific collaboration networks. // *PNAS*. 98, 2(2001), pp. 404-409. DOI: 10.1073/pnas.98.2.404
- [8] Sheffi, Y. *Urban transportation networks: Equilibrium analysis with mathematical programming methods*. Prentice-Hall, Inc. Englewood Cliffs, NJ, 1985.
- [9] Singh, J. *External collaboration, social networks and knowledge creation: evidence from scientific publications*. Danish Research Unit of Industrial Dynamics Summer Conference, Denmark, 2007.
- [10] Tamashita, Y.; Okubo, Y. Patterns of scientific collaboration between Japan and France: Inter-sectoral analysis using Probabilistic Partnership Index (PPI). // *Scientometrics*. 68, 2(2006), pp. 303-324. DOI: 10.1007/s11192-006-0105-1
- [11] Tang, J.; Zhang, J.; Yao, L. et al. ArnetMiner: Extraction and mining of academic social networks. // *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, pp. 990-998. DOI: 10.1145/1401890.1402008
- [12] Wagner, C. S.; Leydesdorff, L. Mapping global science using international co-authorships: A comparison of 1990 and 2000. // *International Journal of Technology and Globalization*. 2, 1(2005), pp. 185-208. DOI: 10.1504/IJTG.2005.007050
- [13] Yan, E.; Ding, Y.; Zhu, Q. Mapping library and information science in China: a coauthorship network analysis. // *Scientometrics*. 83, 1(2010), pp. 115-131. DOI: 10.1007/s11192-009-0027-9
- [14] Yoshikane, F.; Kageura, K. Comparative analysis of coauthorship networks of different domains: the growth and change of networks. // *Scientometrics*. 60, 3(2004), pp. 435-446. DOI: 10.1023/B:SCIE.0000034385.05897.46
- [15] Zitt, M.; Bassecouard, E.; Okubo, Y. Shadows of the past in international cooperation: collaboration profiles of the top five producers of science. // *Scientometrics*. 47, 3(2004), pp. 627-657. DOI: 10.1023/A:1005632319799
- [16] Xi, Y.; Tang, F. Research of organizational multiplex multi-kernel network model. // *Journal of Xi'an Jiaotong University*. 36, 4(2002), pp. 430-435.
- [17] Xiao, Q. A Method for Measuring Node Importance in Hypernetwork Model. // *Research Journal of Applied Sciences, Engineering and Technology*. 5, 2(2013), pp. 568-573.

### Author's address

#### Quan Xiao

1) School of Information Technology, Jiangxi University of Finance and Economics,  
 2) Jiangxi Key Laboratory of Data and Knowledge Engineering, Yuping West Street 556#, (Mailuyuan District), Nanchang, Jiangxi, China, 330032  
 E-mail: xiaoquan@foxmail.com