

Research on the cooperative network of relativity and quantum cosmology researchers

Haiqing Bai¹, Changkai Hu²

¹*School of Computer Science & Engineering Artificial Intelligence, Wuhan Institute of Technology, Wuhan, China*

²*College of Mathematics and Physics, Chengdu University of Technology, Chengdu, China*

Abstract: *The collaboration networks are classified and studied by analysing various properties of the collaboration data of contributing scholars, and the modules with high information flow are identified for identifying the collaboration patterns of authors in science and technology disciplines. The basic model of the collaboration network is described in terms of statistical parameters such as network degree distribution, clustering coefficients, efficiency and average network path length, and then more efficient core communities of the original network are visualized and analysed after k-core decomposition and modularly identified. The present study shows that authors contributing to the field of general relativity and quantum cosmology tend to cooperate with the same fraction of the population, which results in information transfer efficiency in and between core communities being much greater than the overall network efficiency. In addition, there are also individual authors who prefer to do their research independently.*

Keywords: *social network analysis, k-core decomposition, rich-club characteristics, nature of community*

1. Introduction

Nowadays, with the advent of big data and advances in internet technology, information society is evolving at a faster pace. When conducting research activities, researchers are more likely to work in teams and discuss collectively to develop theories and applications more efficiently. Whether in science and technology or in literature, history and the arts, the flourishing and novel breakthroughs require the assistance of scholars from different disciplines, the exchange of ideas and the unity of effort.

It is clear that collaborative networks of scholars are of great importance to research. And therefore various network analysis methods can be used for scholar evaluation studies. For example, community detection is a powerful method which many researchers in this field make numerous innovations^[1,2], and it's useful for various kinds of complex network research. So far, scholars all over the world have made preliminary studies on the different properties of various types of complex networks. In the field of collaborative networks, Yao Kang studied data from patent collaboration networks in the Internet of things technology^[3], and Zhang Yuhao^[4] et al. studied the core members, research teams, and research hotspots in the field of scientific authorship collaborative networks in the field of sports science in China. The study referred to various overall network structure indicators such as network density and centrality, and used social network analysis, bibliometrics and visual mapping to derive recommendations for disciplinary development. In the field of tourism science, Jiang Guangxiu, Li Yongquan^[5] et al. explored the factors influencing innovation performance in terms of collaborative network status, number of relationships, quality of relationships and knowledge diversity, which were analysed through social network analysis and hierarchical regression analysis. In addition, Xuan Liu^[6] et al. investigated the transferability and distribution of research collaboration networks using a random graph model. Because most of the relevant studies are lacking in the collaborative network of science subjects, and the science subjects themselves have the characteristics of complicated branches, it is not easy to conduct a generalized study. This study is based on the submission data of the General Relativity and Quantum Cosmology category in e-print Arxiv^[7] based on the basic topological properties such as degree distributions, clustering coefficients, the shortest path length distributions, and rich-club properties, association properties and densities, to further understand their cooperation patterns.

2. Statistical parameters of the cooperation network

In order to apply complex network theory for research and analysis, this paper abstracts the cooperative network structure based on the Spacer-L network topology model as an unweighted undirected graph $G=\langle V,E \rangle$ consisting of a point set V and an edge set E , represented by an adjacency matrix. The order of the point set is denoted as N . The different authors are defined as nodes and the nodes are represented by numbers. If author i co-authored a paper with author j , the graph G contains undirected edges from i to j . If the paper is co-authored by k authors, a fully connected subgraph of G is generated at k nodes.

Because the data represented in Graph G is large and rich in macroscopic properties, some statistical parameters are introduced to further characterize the collaborative network of scholars as follows.

2.1 Degree, degree distribution and degree correlation of nodes

For an undirected graph $G=\langle V,E \rangle$, $\forall v \in V$, the number of times a node v is an endpoint of an edge is said to be the degree, or degree for short, and is denoted $d(v)$. The degree m_i of the node corresponding to researcher i is equal to the number of times i co-authored with others, and the original data is simplified for the purpose of subsequent studies, so that m_i is the number of collaborators. Let a_{ij} be the matrix element of the adjacency matrix of the collaborative network, i.e. we have

$$a_{ij} = a_{ji} = \begin{cases} 1, & \text{Node } i \text{ connected with Node } j \text{ directly} \\ 0, & \text{Other} \end{cases} \quad (1)$$

The degree of a node is denoted as $k_i = \sum a_{ij}$. The average degree of a network is the average of the degree values k_i of all nodes in the network and is expressed as $\langle k \rangle = \sum k_i / N$. In a cooperative network, the average degree of the network reflects the average number of cooperators. The degree distribution $P(k)$ of an undirected network is the probability that a randomly selected node in the network has degree h , i.e. $p(k) = \sum \delta(k-k_i) / N$. If the degree distribution of the network nodes can be described more accurately by a power rate function, the network is scale-free.

The degree correlation of a complex network reflects the average degree distribution of the nodes with degree k and their neighbours[8], Eq:

$$k_{nn}(k) = \langle k_{nn,i} \rangle = \frac{1}{N'} \sum_{i \in M_k} (k_{nn,i}) = \frac{1}{N'} \sum_{i \in M_k} \left(\frac{1}{k_i} \sum_{j \in N_i} a_{ij} k_j \right) \quad (2)$$

Where N_i denotes the set of i -neighbours of a node, M_k denotes the set of nodes of degree k , and N' denotes the number of nodes of degree k . If there is $d(k_{nn}(k))/dk > 0$, then the network satisfaction is considered positively correlated; if $d(k_{nn}(k))/dk < 0$, then the network satisfaction is considered negatively correlated; if $d(k_{nn}(k))/dk = 0$, then the network satisfaction is considered irrelevant. At a macro level, if the network is positively correlated, also known as congruent, the probability of a node with a smaller degree being connected to a node with a smaller degree and a node with a larger degree being connected to a node with a larger degree in the network is much greater than the probability of other connections occurring; if the network is negatively correlated, also known as heterogeneous, the probability of a node with a smaller degree being connected to a node with a larger degree in the network is much greater than the probability of other connections occurring. Degree uncorrelated can be thought of as a network in which any two nodes have the same probability of being connected, and the degree of any node is not greatly influenced by the surrounding degree values.

2.2 Clustering coefficients of nodes

The clustering coefficient (also known as the cluster coefficient) is defined as the number of edges that actually exist between all neighbouring nodes of node i divided by the theoretical maximum number of contiguous edges, while the clustering coefficient can also be described in terms of the ratio of triangles (complete subgraphs of three nodes) to triads (triangles with one side missing) in the network, giving the defining equation[8].

$$C_i = \frac{2l_i}{k_i(k_i-1)} = \frac{N\Delta}{N_3} \quad (3)$$

Where the number of edges existing between k_i neighbouring nodes of node i is denoted as E_i , $N\Delta = \sum a_{ij} a_{jk} a_{ki}$ ($k > i > j$) denotes the total number of triples containing node i ; $N_3 = \sum (a_{ij} a_{kk} + a_{jj} a_{kk} + a_{kj} a_{kk})$

ak_j ($k > i > j$) denotes the total number of triples containing node i , a_{ij} is the adjacency matrix element.

In a collaborative network of scholars, the clustering coefficient of a node reflects the degree of direct collaboration of that scholar, while the average clustering coefficient C of the network is referred to as the clustering coefficient of the whole network.

2.3 Shortest path length, average path length and efficiency between nodes

In an unweighted network, the minimum number of connected edges that exist between node i and node j is noted as the shortest path length, also known as the geodesic, distance, and is denoted by the symbol d_{ij} . Fig. 1. shows the distribution of path lengths in the collaborative network of scholars in this study. The average path length is then the average of the distances of all nodes in the network, i.e. $\langle l \rangle = \sum d_{ij} / (N(N-1))$ ($i \neq j=1$).

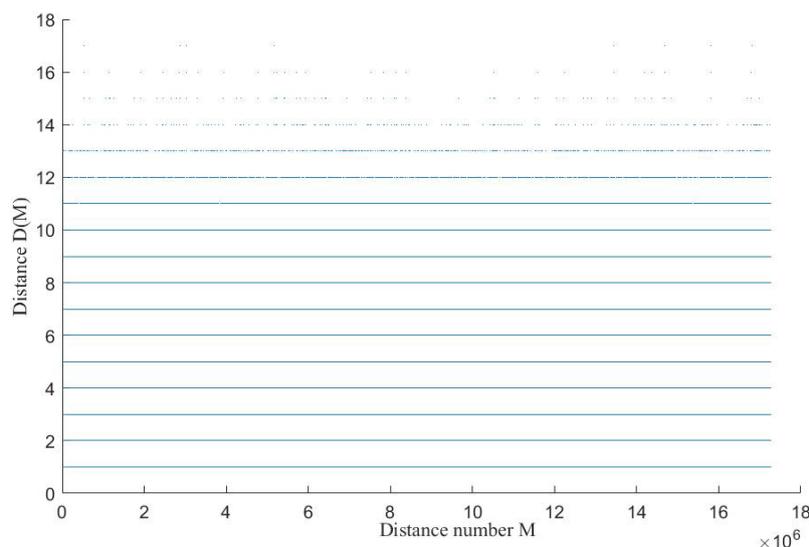


Figure 1: Path Length Distribution

Network efficiency, also known as arrival rate, is expressed as $E = \sum (1/d_{ij}) / [N(N-1)]$ ($i \neq j=1$). In a collaborative network of scholars, the shortest path length represents how closely two scholars collaborate with each other, and efficiency represents how easily scholars can exchange information with each other.

3. Analysis of the structure of the cooperation network

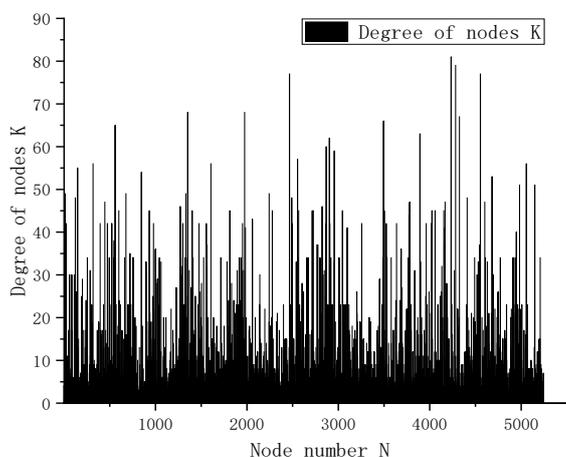


Figure 2: Degree Distribution

By using matlab to calculate and analyse the statistics of the scholar collaboration network, the characteristics of scholar collaboration in this network can be obtained.

First, the network of general relativity and quantum cosmological cooperation is already relatively mature and has a distinctly scale-free nature. As shown in Fig. 2. and Fig. 3, most of the node degrees are distributed between 0 and 20. From Fig. 4, the cumulative degree distribution of the network logarithmically satisfies $y=kx+b$ (calculated by fitting $k=-2.042$, $b = 0.2312$), which fits the power law distribution.

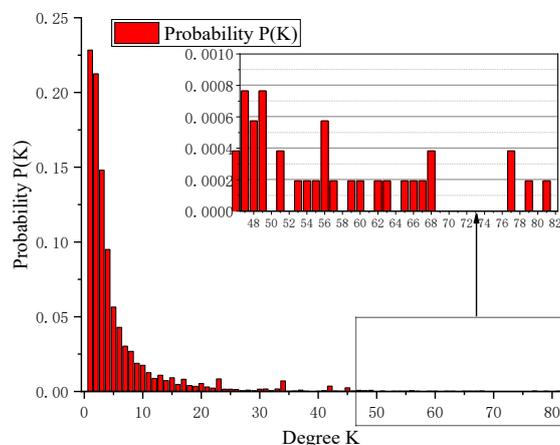


Figure 3: Cumulative Degree Distribution

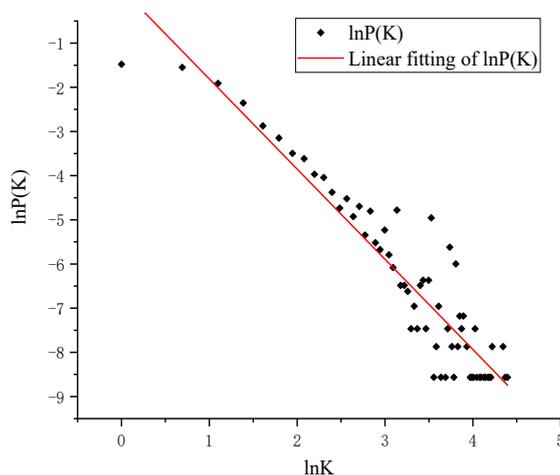


Figure 4: Cumulative Log Diagram

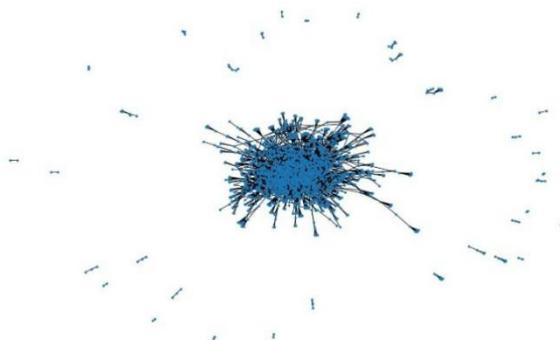


Figure 5: Cooperation Network

Second, the network of general relativity and quantum cosmological cooperation exhibits certain small-world effects. The overall network is shown in Fig. 5. If a network is said to have a small-world effect, it means that information is transmitted more smoothly in the network and is less prone to information lag. The small-world nature of the network can be described by the characteristic path length and the clustering coefficient. A large clustering coefficient represents a strong aggregation across the network, while a longer characteristic path length indicates that it is more difficult for scholars to establish collaborative relationships. As can be seen from the data in Table 1, the average distance across

the collaborative network is short and has a high clustering coefficient with a strong small-world effect.

Table 1: Statistical parameters of original cooperative network

| Node | Average Degree | Clustering coefficient | Network Characteristic path length | Efficiency | Density | The Largest Degree | The Minimum Degree |
|------|----------------|------------------------|------------------------------------|------------|---------|--------------------|--------------------|
| 5242 | 5.5284 | 0.5296 | 6.0485 | 0.1793 | 0.0011 | 108 | 1 |

The sheer size of collaborative networks of papers spanning a large period of time and within a single discipline does not easily lend itself to characterisation to analyse the specific nature of the network.

3.1 Analysis of the community structure

Due to some objective reasons, the scholars' thesis cooperation network has a more obvious association structure. Within those communities, the flow of information is more frequent.

The relationship between the degree of nodes and connectivity allows us to determine the extent to which communities contribute to the total network information. In order to highlight the communities constituted by nodes with frequent information transmission, the concept of k-core community is proposed in this paper. Its formation is first obtained by k-core decomposition of the corresponding sub-networks, and then the sub-networks are divided to different core communities by a recorded algorithm[9], the results are shown in Fig. 6, Fig. 7. and Fig. 8.

By comparing the values recorded in Table 2, it is clear that the original network is modular and core communities have shorter path lengths and more efficient information transfer than the original network.

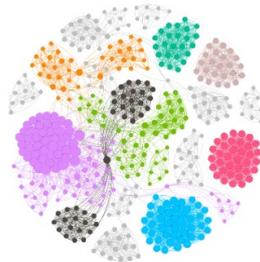


Figure 6: Network 1

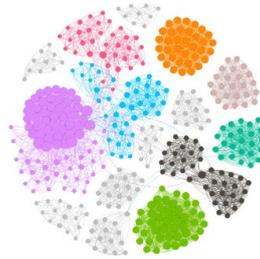


Figure 7: Network 2

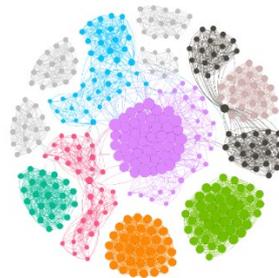


Figure 8: Network 3

Table 2: Comparison of statistical parameters of cooperative network

| Name | Modularity | Number of Communities | Efficiency | Network characteristic path length |
|------------------|------------|-----------------------|------------|------------------------------------|
| Original Network | 0.859 | 387 | 0.1793 | 6.0485 |
| Network1 | 0.840 | 18 | 0.3301 | 4.6164 |
| Network2 | 0.837 | 15 | 0.4471 | 3.0019 |
| Network3 | 0.833 | 12 | 0.4843 | 2.7410 |

3.2 Rich-Club Characterisation

Rich nodes are a few nodes in a graph with a large number of edges, and the rich-club property is the characteristic that rich nodes tend to be connected to other rich nodes rather than non-rich nodes, and the coefficient used to measure this property is known as the rich-club property, Eq.[10] Written by

$$\Phi(r/N) = \Phi(k) = \frac{L}{r(r-1)/2} = \frac{2L}{r(r-1)} \quad (4)$$

For the collaborative network as a whole, the change in the rich-club coefficient with the value of k is shown in Fig. 9. As can be seen from the figure, the rich-club coefficient increases with increasing values of k. When the value of k is greater than 15, the rich-club coefficient is greater than 0.6.

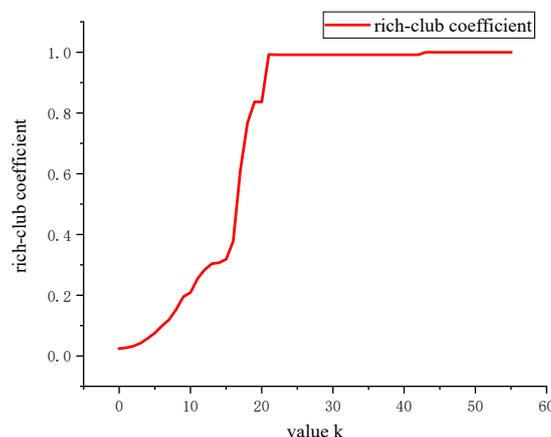


Figure 9: Change of rich-club coefficient

3.3 Heterozygosity, mediator

The congruence of the network can be described by the congruence coefficient equation[10] as follows

$$r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2} \quad (5)$$

Table 1 shows that the coefficient of congruence of the collaborative network is greater than zero and the network is congruent, which again demonstrates that academics who are happy to collaborate are more likely to collaborate with other academics who are more collaborative.

Medial centrality can characterize a node's control over the transmission of information along the shortest path between pairs of nodes in a network, Equation[11] is expressed as $BC_i = \sum_{st} n_{ist} / g_{st}$, where g_{st} is the number of shortest paths from node s to node t and n_{ist} is the number of shortest paths from node s to node t through node i. Values of betweenness centrality of nodes are shown in Fig. 10.

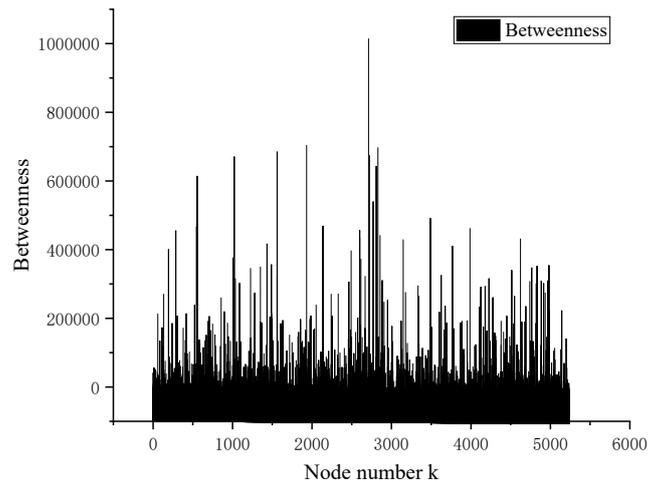


Figure 10: Betweenness centrality

4. Conclusion

This paper innovates the concept of k -core community, which is then used to uncover the more efficient parts of larger collaborative network, and synthesizes the model that scholars in the field prefer to conduct their research. This paper uses the Space-L network topology model as the basis for computing and analysing the statistical parameters of this cooperative network, and then further analyse the overall network and certain characteristics of the sub-networks in order to gain a deeper understanding of the model of cooperation between scholars in this reserch field, which better promote scientific cooperation among scholars.

Research has shown that there are two types of scholars in this field: those who are willing to collaborate and those who tend to work independently. The willing authors tend to collaborate indirectly or directly with a fixed group of people, and there are dozens of core communities in the large collaborative network of willing scholars based on cooperative relationships, and the efficiency of information transfer within and outside the core communities is much greater than that of the network as a whole.

However, there are still some limitations in this paper. Firstly, this study does not distinguish between first and second works, as first works contribute more than second works and should be given a greater weighting in the information flow network; secondly, this study only considers one network of author research collaboration within the field and fails to consider the interaction between multiple networks, such as the literature citation network, which is non-ignorable for relavent research.

References

- [1] Ye Wenbin. *Research on community discovery algorithm of spectral clustering based on Bayesian inference and graph kernel [D]*. South China University of Technology, 2021. DOI: 10.27151/d. CNKI. Ghnl. 2021.002206.
- [2] Wang Min. *Research on key node mining and community discovery algorithms in complex networks [D]*. University of Electronic Technology, 2020. DOI: 10.27005/d. CNKI. GDZKU.2020.001854.
- [3] Yao Kang. *Research on the impact of patent cooperation network in IOT technology on innovation performance [D]*. University of Electronic Technology, 2022.
- [4] Zhang Yuhao, Jiang Chao, Wang Peng, Wu Youliang, Zhou Haikui. *Analysis of collaborative networks and collaborative research hotspots of sports science authors in China [J]*. Journal of Liupanshui Normal College, 2021, 33(05):109-120. DOI: 10.16595/j.1671-055X.2021.05.013.
- [5] Jiang Guangxiu, Li Yongquan, Ruan Wenqi, Zhang Shuning, Xu Anan. *A study on the mechanism of influence of tourism scholars' cooperative network status on innovation performance [J/OL]*. Journal of Tourism: 1-21[2022-08-22]. DOI: 10.19765/j.cnki.1002-5006.2022.00.007.
- [6] Liu Xuan, Wang Linwei, Li Jia, Zhang Pengzhu *Mechanism of scientific research cooperation network formation--an analysis based on stochastic exponential graph model [J]*. Journal of Systems Management, 2019, 28(03):520-527.

- [7] J. Leskovec, J. Kleinberg and C. Faloutsos. *Graph Evolution: Densification and Shrinking Diameters*. *ACM Transactions on Knowledge Discovery from Data (ACM TKDD)*, 1(1), 2007.
- [8] He Daren, Liu Zonghua, Wang Binghong. *Complex networks and complex systems [M]*. Beijing: Higher Education Press, 2009, 126-157.
- [9] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, *Fast unfolding of communities in large networks*, in *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10), P1000.
- [10] Wang Xiaofan, Li Xiang, Chen Guanrong *Complex network theory and its applications [M]* Tsinghua University Press 2006 55-56.
- [11] Wang Xiaofan, Li Xiang, Chen Guanrong *Introduction to Network Science [M]* Beijing: Higher Education Press 2012 P159.