Chaos, Solitons and Fractals 114 (2018) 461–467

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

The agglomeration phenomenon influence on the scaling law of the scientific collaboration system

Ai-Zhong Shen[a], Jin-Li Guo[a,b,*], Guo-Lin Wu[a], Shu-Wei Jia[a]

[a] Business School, University of Shanghai for Science and Technology, Shanghai 200093, China
[b] Center for Supernetwork Research, University of Shanghai for Science and Technology, Shanghai 200093, China

ARTICLE INFO

Article history:
Received 3 January 2018
Revised 24 July 2018
Accepted 30 July 2018

Keywords:
Scientific collaboration
Hypergraph
Hypernetwork
Scaling law
Clustering coefficient

ABSTRACT

This paper presents a scientific collaboration hypernetwork evolution model with adjustable clustering coefficient in which the authors are regarded as nodes and the hyperedges are the cooperative articles. Firstly, we build the scientific cooperation hypernetwork through the real paper data which comes from the database of arxive. The empirical results show that the node's hyperdegree follows the power law distribution, but the hyperedge's node-degree has the exponent distribution. In addition, we establish the scientific collaboration clustering hypernetwork evolution model by employing the Poisson process theory and the continuous method for studying the agglomeration phenomenon of scientific cooperation system. The theoretical analysis shows that the node's hyperdegree distribution has scale-free characteristics. The power index of our model is independent of the clustering coefficient, and the theoretical analyses agree with the conducted numerical simulations. Moreover, our model not only describes the scale-free property, but also depicts the phenomenon of agglomeration. Both the properties are usually coexist in the scientific cooperation, our model is more actual.

1. Introduction

Complex networks theory [1–3] is an important tool for studying the evolution and the statistical characteristics of the real system. After more than ten years of development, the theory has been widely applied to various scientific field [4–7], such as politics, economy, society and biology, etc. In the complex networks, nodes correspond to different individuals involved in the study, while edges between nodes represent relationships between the connected nodes in the real system. However, the theory of complex networks portrays the real system also have some limitations and lose a lot of information of systems. For example, each edge connects only two nodes. The relationships among the objects of each complex real system tend to be more complex than those that can be described with simple pairwise relations. For example, the urban rail transit system which not only exist the network of the site relationships [8], but also have the power network [9]. It is necessary to expand the traditional complex network theory to ensure the information integrality and express the relationships among more than two nodes in the real system. Recently, the concept of hypernetworks [10] has been attracted much attention in the scientific community.

The cooperation exist in society widely, such as the actor cooperation networks [11], the enterprise cooperation networks [12], and the scientific cooperation networks [13]. Some works studied the relationships about the scientific cooperation with the complex networks. In the scientific cooperation complex network, nodes represent the authors and the edges are the cooperation relationships among these authors. The works of scientific cooperation with complex network are mainly in the empirical analysis and the evolution model. Newman et al. [13–16] analyzed empirically the scientific cooperation networks and calculated the statistical indicators through different data. Sarkar et al. [17] studied the collaboration network under the random matrix theory framework. The results showed that the collaboration network involves a large number of complete subgraphs and the collaboration groups among scientists. Barabâsi et al. [18] analyzed the data of neural and mathematical disciplines, and established the evolution model of the complex network. The result showed that the complex network has the scale-free property. Zhou et al. [19] proposed a more general networks model in which the degree distribution is between the exponential and power-law distribution. Li *et al.* [20] proposed evolution model of weighted networks which is inspired by empirical analysis the data of econophysicists. The model introduced three mechanisms: the old vertices can build up

* Corresponding author at: Business School, University of Shanghai for Science and Technology, Shanghai 200093, China.
E-mail addresses: phd5816@163.com, 275455466@qq.com (J.-L. Guo).

https://doi.org/10.1016/j.chaos.2018.07.037

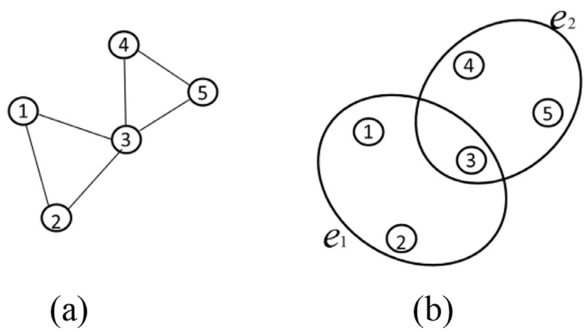

Fig. 1. Compare the complex network and the hypernetwork to depict the scientific cooperation system. Fig. 1 represents the cooperation relationships among the five authors. (a) the complex network depicts the cooperation relationship which show the author's relationship, but cannot get the author's article information. (b) the hypernetwork depicts the cooperation relationship in which the closed curve is the hyperedge and the nodes inside a hyperedge are connected fully. It can get the information about the cooperation relationship and the number of the articles of any author. For example, the author 3 cooperates with the authors 1,2,4,5 to write two articles e_1, e_2, the node's hyperdegree of the node 3 is 2.

new links, the evolution of link weight and the referential attachment by the local information. Viana et al. [21–22] calculated some statistical of the scientific cooperation networks, such as the degree distribution, the network diameter and the network density. Ramasco [23] proposed the collaboration network in terms of the evolution and self-organization bipartite graph model.

Most the studies of scientific cooperation relationships above are based on the complex network theory. However, complex networks only study the cooperation relationships among the authors. The information about the number of author's articles and an article contains the number of authors be lost, therefore we are not aware of the author's actual contribution of scientific research in a way. Comparing with the complex network theory (Fig. 1), the hypernetwork theory more adequately describes the scientific cooperation relationships. Hu et al. [24] established the scientific cooperation hypernetworks model which involves the basic features of BA model [2], the growth and preferential attachment. Yang et al. [25–26] built local preferential hypernetworks model of scientific cooperation in which the new nodes only can select local old nodes to form hyperedges. Li et al. [27] combined two preferential mechanism, the different way of knowledge growth and the knowledge diffusion, to construct the knowledge diffusion hypernetworks evolution model.

However, the common feature of the earlier works of the hypernetwork model [24–27] reviewed above is that the evolution mechanism is an extension of complex networks. In order to establish an evolution model conforms to the reality of the evolution law of scientific collaboration, the different hypernetwork evolution model is proposed in this paper. In reality, there are three evolution mechanisms of the scientific cooperation hypernetwork in this paper. *Firstly*, the scale of the hypernetwork grow constantly. The earlier works [13–27] about the scientific cooperation model assumed that the nodes enter the hypernetwork at the equal time interval, it is not reasonable for the reality of the hypernetwork growth. In this paper, we take nodes arrive into the hypernetwork as an event, the Poisson process can better describe the growth of the hypernetwork. *In addition*, the new nodes preferentially attach mechanism. Most of the previous collaboration hypernetworks [24–27] are based on the complex networks theory in which the new node's preferential attachment is based on the node degree. It means that the more the number of author's collaborators, the easier the author is connected. In reality, scholars tend to select the collaborators who have more number of the scientific papers which can show the author's scientific contribution to some extent. The node's hyperdegree can show the number of author's papers in the hypernetwork theory. In this paper, the new node's preferential attachment is based on the node's hyperdegree. It means that the more the number of author's papers, the easier the node is connected by new nodes. The model can better depict the facts of the new node's preferential attachment mechanism. *Moreover*, the agglomeration phenomenon is common in social systems and the cooperative relationship involves the clustering phenomena in the scientific research. Scholars who are the same discipline or research team are more likely to be chosen as a collaborator by author, which lead to a large number of clustering phenomena in the scientific cooperation. However, most of the earlier hypernetwork model [24–32] are based on the BA model [2] which has a smaller clustering coefficient. There are no hypernetwork evolution models with larger clustering coefficient and the existing models only depict the power-law characteristics. In this paper, we build the model involves the clustering attachment mechanism which is able to capture the phenomenon of clustering in the scientific cooperation.

The rest of the paper is organized as follows. In next section, we introduce respectively the concepts and statistical indicators of scientific cooperation hypernetworks. After that, we empirically study the scientific cooperation date which come from the database of arxive. That is followed by the theoretical analysis and numerical simulations of the scientific collaboration hypernetwork evolution model with adjustable clustering coefficient .At the end, our presentation is concluded with some conclusion remarks.

2. The related concepts

The hypernetworks can be categorized into those of network-based supernetworks [33] and hypergraph-based hypernetworks [34]. This paper mainly researches the evolution of the hypergraph-based hypernetworks which are a dynamic hypergraph in which edges are regard as hyperdeges, each hyperedge contains arbitrary number of nodes. Ref. [35] gives the mathematical definition of hypergraph-based hypernetworks. $\Omega = \{(V, E^h)|(V, E^h)\}$ is a finite hypergraph and G^h is a mapping from $[0, +\infty)$ to Ω, where $G^h(t) = (V(t), E^h(t))$ is a finite hypergraph for any given $t \geq 0$. Here the indicator t is often interpreted as time. We consider that $G^h(t) = (V(t), E^h(t))$ is a hypernetwork for sufficiently large t, where $V(t) = \{v_1, v_2 \cdots v_t\}$ indicates the set of nodes and $E^h(t) = \{e_1, e_2 \cdots e_t\}$ indicates the set of hyperedges. In this paper, the related statistical indicators are defined as follows:

(1) The hyperdegree of nodes. The hyperdegree k_i^h of node v_i is defined as the number of hyperedges contain the node. In the Fig. 1(b), the hyperdegree of nodes from v_1 to v_5 are 1, 1, 2, 1, and 1 respectively.
(2) The node-degree of hyperedges. It refers to the number of nodes in a hyperedge. It means the number of authors in a paper in the scientific collaboration hypernetwork.
(3) The hyperdegree distribution of nodes. The hyperdegree distribution of nodes is defined as $P_i(t, k) = P\{k_i^h(t) = k\}(k = 1, 2, \cdots)$, where $k_i^h(t)$ denotes the transient hyperdegree of node i at time t.$\{P_i(k)\}$ denotes the stationary hyperedegree distribution of node i, where $\sum_{k=0}^{\infty} P_i(k) = 1$ and $\lim_{t\to\infty} P\{k_i^h(t) = k\} = P_i(k)(k = 1, 2, \cdots)$ are existence.
(4) The hyperdegree distribution of the hypernetwork. The hyperdegree distribution of hypernetworks is defined as $P(t, k) = \frac{1}{n(t)} \sum_{i\in V(t)} P\{k_i^h(t) = k\}$, where $n(t)$ denotes the total number of nodes in the hypernetwork at time t. $P(k)$ denotes the stationary hyperdegree distribution of the hypernetwork, where $\sum_{k=0}^{\infty} P(k) = 1$ and $\lim_{t\to\infty} P(t, k) = P(k)$ are existence.