

# Structure Analysis in Hypernetwork of Human Protein Complexes

XiMei Hou<sup>1,2\*</sup>, GaoXia Wang<sup>1,2</sup>

<sup>1</sup>(College of Science, China Three Gorges University, China)

<sup>2</sup>(Mathematics Research Center, China Three Gorges University, China)

---

**Abstract:** In this paper, a hypernetwork model is constructed by taking human proteins as nodes and protein complexes formed by protein interactions as hyperedges. By calculating and analyzing the related indexes of the hypernetwork, it is found that it has the nature of small world, and the node's hyperdegree presents the scale-free characteristic, but its hyperedge hyperdegree does not meet the scale-free characteristic. Considering the cross-linking of three hyperedges in the hypernetwork, the hypergraph motif is used to analyze the structure of the human protein complex hypernetwork.

**Key Word:** Hypernetwork; Protein complex; Topological property analysis; Hypergraph motifs.

---

Date of Submission: 12-02-2023

Date of Acceptance: 24-02-2023

---

## I. Introduction

In the real society, the research objects are usually regarded as nodes, and the relationships between node pairs are mapped to edges to build a complex network for research. However, the relationship between node pairs cannot reflect the relationship between multiple nodes. For example, the relationship formed by the node pairs cannot reflect the characteristics of multiple chemicals participating in the metabolic reaction in the metabolic process of the organism. Therefore, the hypernetwork based on hypergraph is introduced to better represent the relationship between multiple nodes.

With the development of society, win-win cooperation becomes a trend, and people tend to cooperate among groups. In order to reflect the characteristics of inter-group connections, Berge proposes the concept of hypergraph<sup>[1]</sup>, that is, hypergraph is a graph used to represent the relationship between multiple nodes. The network based on hypergraph structure is called hypernetwork<sup>[2]</sup>.

In the traffic network, literature [3-5] study the topological characteristics and robustness of aviation and public transport hypernetworks, and attempt to build urban traffic network models with strong robust topology. In the social network, the hypernetwork models of user association under social software<sup>[6]</sup> and product competitiveness analysis<sup>[7]</sup>, the knowledge diffusion model of scientific research cooperation hypernetwork<sup>[8]</sup>, and the word formation rule model of Tang poetry and Song poetry<sup>[9]</sup> are constructed based on the hypergraph.

In the biological network, Percy et al.<sup>[10]</sup> map the chemical substances existing in the organism as nodes, and the chemical reactions jointly participated by nodes are regarded as hyperedges to study the complexity and robustness of the metabolic hypernetwork. Since a single protein can participate in the formation of multiple complexes and a complex may contain multiple proteins, Ramadan et al.<sup>[11]</sup> study the relationship between multiple proteins in yeast protein complex based on the hypernetwork. Gouchol et al.<sup>[12]</sup> construct a small world model of protein complex network and study the centrality of nodes and the community structure of the network. Hu Feng<sup>[13]</sup> identifies key proteins of the protein complex hypernetwork through related topological indexes.

In this paper, proteins are regarded as nodes and complexes formed by protein interactions are regarded as hyperedges based on the definition of protein complex hypernetwork proposed by Ramadan et al.. The data about human proteins forming complexes are collected from the database to analyze and study the characteristics and structure of human protein complex hypernetwork. Section 2 introduces the concept of hypergraph and related topological indexes. Section 3 introduces the data source of human protein complexes and analyzes the related topological properties of human protein complex hypernetwork. Section 4 summarizes the conclusions.

## II. Preliminaries

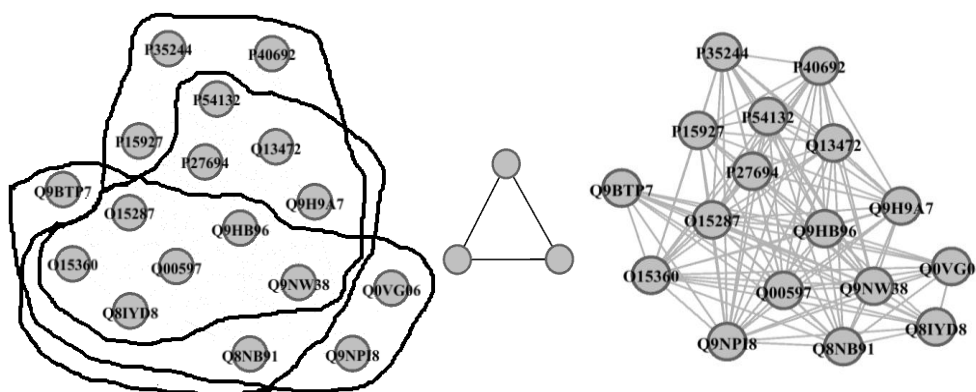
### 1. Representation of hypergraph

#### (1) General representation of hypergraph

Hypergraph<sup>[1]</sup> refers to a binary relationship  $H = (V, E)$ , where the elements of set  $V$  are the nodes of the hypergraph, represented by  $V = \{V_1, V_2, \dots, V_N\}$  and the elements of set  $E$  are the hyperedges of the hypergraph, represented by  $e_i = \{V_{i1}, V_{i2}, \dots, V_{ir}\} (r \leq N)$  satisfying  $e_i \neq \Phi (i = 1, 2, \dots, M)$ ,  $\bigcup_{i=1}^M e_i = V$ .

The edges of a normal graph represent the connections between node pairs. The hyperedges can contain any number of nodes and represent the connections of any node group.

Hypergraphs are obtained by using nodes to represent the elements in  $V$  and closed curves containing nodes to show that there is a certain relationship between multiple nodes. Figure 1a shows the composition of three human protein complexes.



(a)Hypergraph (b)Projected graph(c)2-section graph

**Figure no 1:** Hypergraph and its transformation

#### (2) Projected graph of hypergraph

The projected graph<sup>[14]</sup> of the hypergraph is defined as  $\bar{H} = (E, \wedge)$ . The hyperedges in  $H$  are regarded as the nodes in the projected graph. When there are the common nodes between two hyperedges, the nodes representing the hyperedges in the projected graph are connected with the edges  $\wedge$ .  $\wedge_{ij}$  represents the edge

formed by hyperedges  $e_i$  and  $e_j$  in the projected graph, expressed as  $\wedge_{ij} = \left\{ (e_i, e_j) \mid (e_i, e_j) \in \binom{E}{2}, e_i \cap e_j \neq \Phi \right\}$ .

Figure 1b is the projected graph of the hypergraph corresponding to Figure 1a.

#### (3) 2-section graph of hypergraph

The nodes of 2-section graph  $\tilde{H}$  are the nodes of the hypergraph  $H$ . When two nodes in  $H$  belong to the same hyperedge, there is an edge between the two nodes. The network formed is called the 2-section graph of the hypergraph<sup>[15]</sup>. Therefore, nodes belonging to the same hyperedge are fully connected in the 2-section graph. Figure 1c is the 2-section graph of the hypergraph corresponding to Figure 1a.

## 2. Topological indicators of the hypernetwork

### (1) Hyperdegree

The hyperdegree of node refers to the number of hyperedges containing node  $V_i$  in the hypernetwork, which is recorded as  $Hd(V_i)$ . In the protein complex hypernetwork, the hyperdegree of node represents the number of protein complexes containing the corresponding node. Hyperdegree  $Hd(e_i)$  of hyperedge refers to the number of nodes contained in hyperedge  $e_i$  in the hypernetwork.

### (2) Clustering coefficient

Clustering coefficient is a parameter that describes the aggregation degree of nodes in general networks. The local clustering coefficient  $C_i$  of node  $i$  is the ratio of the number of connections between its neighboring nodes to the number of all possible connections between them. The clustering coefficient  $\langle C \rangle$  of the

whole network is defined as the average value of the local clustering coefficient of nodes in all connected branches with non-zero local clustering coefficient, expressed as  $\langle C \rangle = \frac{\sum_{\omega_s \in \omega} \sum_{i=1, \dots, |\omega_s|} C_i}{\sum_{\omega_s \in \omega} |\omega_s|}$ , where  $\omega$  represents the set of connected branches in the network, and  $\omega_s$  represents the connected branch with non-zero local clustering coefficient.

The local clustering coefficient of the hypernetwork is expressed as the degree of connection between adjacent hyperedges of a hyperedge. In the hypernetwork, the hyperedges correspond to nodes in the projected network, and the common nodes between the hyperedges correspond to edges in the projected network. Therefore, the calculation of the clustering coefficient of the hypernetwork can be transformed into the calculation of the clustering coefficient of the projected network, which is called the clustering coefficient  $\langle \bar{C} \rangle$  based on the hyperedge, where the local clustering coefficient  $\bar{C}_i$  is expressed as  $\bar{C}_i = \frac{2L_i}{K_i(K_i - 1)}$ , and  $L_i$  is the number of links between  $K_i$  neighbors of node  $i$  in the projected graph.

At the same time, the clustering coefficient of the hypernetwork can also be seen as the probability that any two neighbor nodes of the node are connected. It is equivalent to the clustering coefficient  $\langle \tilde{C} \rangle$  of the 2-section graph, which is called the clustering coefficient based on the node. Its local clustering coefficient  $\tilde{C}_i$  is expressed as  $\tilde{C}_i = \frac{2E_i}{N_i(N_i - 1)}$ , where  $N_i$  is the degree of node  $i$  in the 2-section graph, and  $E_i$  is the number of links between the  $N_i$  neighbor nodes of node  $i$ .

### (3) Average path length

The average path length refers to the average value  $\langle d \rangle$  of the distance between node pairs in all connected branches of the network and is expressed as  $\langle d \rangle = \frac{\sum_{\omega_s \in \omega} \sum_{i \neq j, i, j=1, \dots, |\omega_s|} d_{ij}}{\sum_{\omega_s \in \omega} |\omega_s|(\omega_s - 1)}$ .  $d_{ij}$  and  $d_{ji}$  represent the distance between node pairs  $i$  and  $j$  in connected branch  $\omega_s$ , with  $d_{ij} = d_{ji}$ .

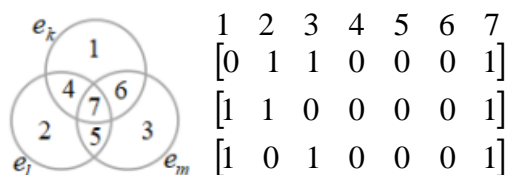
The average path length in the hypernetwork can be defined as the average value of the distance between any two hyperedges in all connected branches. When the hypernetwork is mapped to the projected graph, the distance between the two hyperedges corresponds to the distance between two nodes in the projected network. Therefore the average path length in the projected graph is called the average path length  $\langle \bar{d} \rangle$  of the hypernetwork based on the hyperedge. At the same time, the average path length in the hypernetwork is expressed as the average of the number of hyperedges on the shortest path connecting two nodes, namely, the average path length of the 2-section graph, which is called the average path length  $\langle \tilde{d} \rangle$  of the hypernetwork based on the node.

### 3. Hypergraph motifs

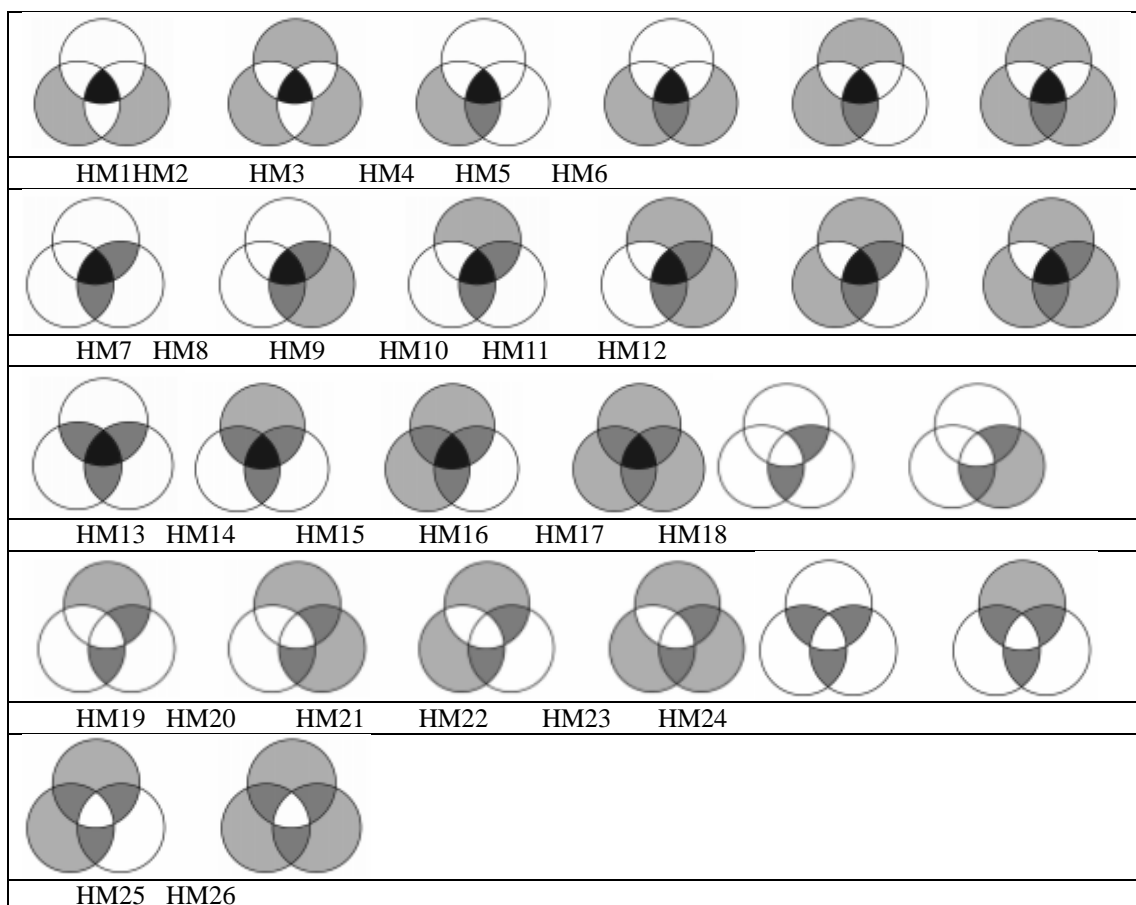
In the general network graph, the network is analyzed through the three-node motifs that can reflect the function of the network<sup>[16]</sup>. Generalized to hypernetwork, Lee et al.<sup>[14]</sup> propose the concept of hypergraph motifs through Venn diagram to explore the relevant structure of the hypernetwork. Here, hypergraph motifs are defined by using vector.

If two hyperedges have common nodes, the two hyperedges are adjacent. When the connectivity of three different hyperedges  $e_k$ ,  $e_l$  and  $e_m$  is considered at the same time, the cases of nodes belonging to three superedges can be divided into 7 types (Figure 2a). There are three cases  $e_k - e_l - e_m$ ,  $e_l - e_k - e_m$  and  $e_m - e_k - e_l$  where the nodes belong to only one hyperedge, three cases  $(e_k \cap e_l) - e_m$ ,  $(e_l \cap e_m) - e_k$  and  $(e_k \cap e_m) - e_l$  where the nodes belong to only two hyperedges, and one case  $e_k \cap e_l \cap e_m$  where the nodes belong to three hyperedges simultaneously. The  $7 \times 1$  order matrix is used to represent the existence of nodes in these 7 parts of any three hyperedges. If nodes exist, the element of the matrix is represented by 1, otherwise, the element is 0 (Figure 2b). Since nodes may exist in any 7 cases, there is  $2^7$  combination modes. Excluding the combination modes with repeated hyperedges and disconnected, there are still 86 combination modes,

constituting 26 interaction cases of three mutually different hyperedges. These 26 cases are called hypergraph motifs (Figure3).



(a)The partition of nodes in hyperedges(b)Vector representation of the same hypergraph motif HM1  
**Figure no 2:** Adjacency of three hyperedges



**Figure no 3:** Hypergraph motifs<sup>[14]</sup>

### III. Result

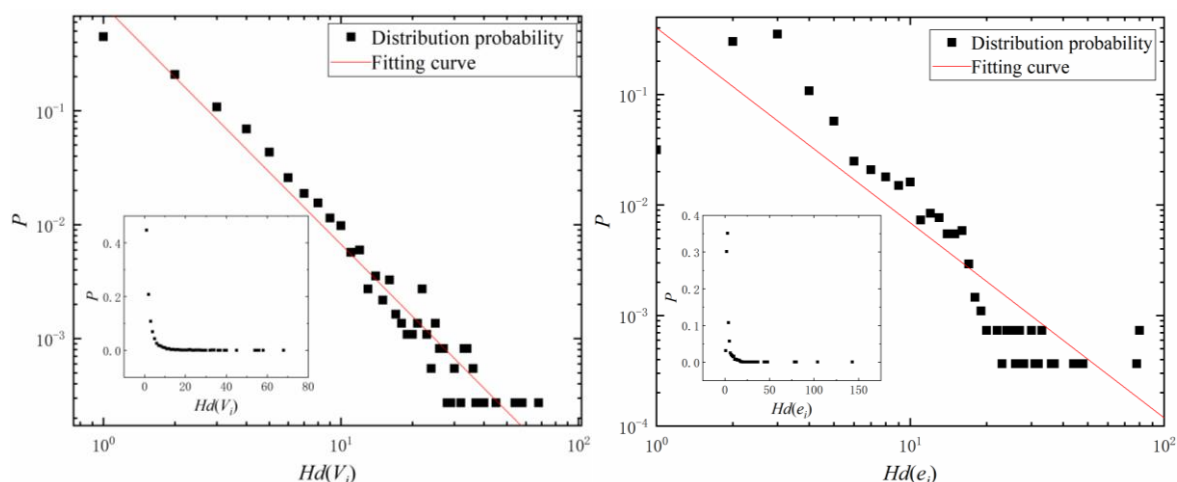
Data on the formation of human protein complexes are collected from CORUM(Comprehensive Resource of Mammalian protein complexes)<sup>[17]</sup>(<http://mips.helmholtz-muenchen.de/corum/>).There are 2734 protein complexes formed by the interactions of 3,674 proteins.Based on the definition of hypergraph, a hypernetwork model of protein complexesis constructed by using relevant data. In this model, nodes represent proteins and hyperedges represent complexes composed of protein interactions.Figure 1a shows the partial composition of the human protein complex hypernetwork.

#### 1. Hyperdegree distributions of nodes and hyperedges

In the protein complex hypernetwork, the hyperdegree of node represents the number of protein complexes containing the corresponding node.The hyperdegree of node in the hypernetwork is plotted using Cartesian coordinates (small figure) and double logarithmic coordinates (large figure) to obtain the hyperdegree distribution of human proteins(Figure 4a).The slope of the fitting curve of the double logarithmic degreedistribution of the nodes in Figure 4a is -2.08, which follows the power law distribution<sup>[18]</sup>. This meansthat the protein complex hypernetwork has scalefree property, that is, only a few proteins participate in the formation of more complexes. And the scalefree property becomes more obvious with the increase of the

number of proteins.

Hyperdegree of hyperedges refers to the number of proteins constituting the protein complexes. Figure 4b shows hyperdegree distribution of human protein complexes in the database. It can be seen that the probability of complexes composed of 2 or 3 proteins is relatively high. The slope of the fitting curve of the double logarithmic degree distribution of the hyperedges in Figure 4b is -1.76, which does not meet the power law distribution.



(a)Hyperdegree distribution of nodes (b)Hyperdegree distribution of hyperedges

**Figure no 4:** Hyperdegree distributions of proteins and protein complexes

## 2. Analysis of average clustering coefficient and average pathlength

For the hypernetwork model representing the relationship between multiple proteins, Matlab and Gephi are used to convert it into the projected graph and 2-section graph. And the average clustering coefficient and average path length are calculated for different network manifestations to characterize the average clustering coefficient and average path length of the hypernetwork. The results are shown in Table 1.

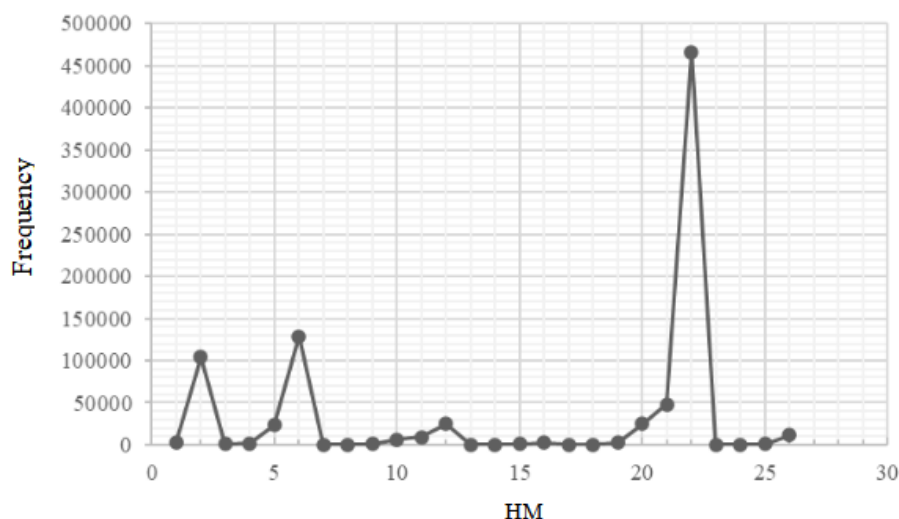
**Table no 1:** Average clustering coefficient and average pathlength

hyperedge based				node based			
$\langle \bar{C} \rangle$	$\langle \bar{d} \rangle$	$\langle k \rangle$	$\ln \hat{N} / \ln \langle k \rangle$	$\langle \tilde{C} \rangle$	$\langle \tilde{d} \rangle$	$\langle k \rangle$	$\ln \hat{N} / \ln \langle k \rangle$
0.723	4.4	21.894	2.564	0.785	4.845	21.536	2.674

It can be seen from Table 1 that the clustering coefficients based on hyperedges and nodes are relatively high, which indicate that the protein complexes in the hypernetwork are closely related. At the same time, the average path length based on the hyperedges and nodes are relatively low, that is, proteins can connect to other proteins through a short path. And proteins are closely related. When the network has high clustering coefficient and low average path length, it is said to have small world property<sup>[18]</sup>.  $\langle d \rangle \approx \ln \hat{N} / \ln \langle k \rangle$  is satisfied, where  $\hat{N}$  represents the number of hyperedges or nodes  $N$  in the hypernetwork, and  $\langle k \rangle$  represents the average degree of nodes in the projected network or 2-section network. It can be seen from Table 1 that the average path length of the hypernetwork satisfies the mathematical expression of the small world property, indicating that the hypernetwork of protein complex has the small world property.

## 3. Motif analysis of hypernetwork

According to the hypernetwork constructed from the human protein complexes in the database, the association of three different hyperedges in the hypernetwork is obtained. And the type and number of hypergraph motifs constituting the hypernetwork are obtained. The results are shown in Figure 5.



**Figure no 5:** Type and number of hypergraph motifs in human protein complex hypernetwork

As can be seen from Figure 5, the hypergraph motifs that constitutes the human protein complexes are obtained. HM22 is the type of hypergraph motifs that appears most frequently.

#### IV. Conclusion

In this paper, the proteins is regarded as the nodes and the protein complexes as the hyperedges based on the database to build a hypernetwork model. By calculating the hyperdegree of node and hyperedge, clustering coefficient and average path length of the hypernetwork, it is found that the hypernetwork is scalefree and satisfies the small world property. Considering the cross-linking of the three complexes in the hypernetwork, the type and number of hypergraph motifs in the hypernetwork are determined, which is conducive to further understanding the structure of the human protein complex hypernetwork.

#### References

- [1]. Berge C. Graphs and hypergraphs[M]. Amsterdam: North-Holland Publishing Company, 1973.
- [2]. Estrada E, Rodríguez-Velázquez JA. Subgraph centrality and clustering in complex hypernetworks[J]. *Physica A: Statistical Mechanics and its Applications*, 2006, 364:581-594.
- [3]. Yonghao S, Jinli G. Topology and robustness analysis of airline hypernetwork[J]. *Intelligent Computer and Applications*, 2021, 11(12): 87-92+96.
- [4]. Haixiu L, Haixing Z, Yuzhi X, et al. A hypergraph-based analysis of the topology and robustness of bus hypernetworks[J]. *Journal of Southwest University (Natural Science Edition)*, 2021, 43(10): 181-191.
- [5]. Ruiming L, Jinli G. Topological characteristics and robustness analysis of shanghai bus hypernetwork[J]. *Mathematics in Practice and Theory*, 2018, 48(20): 129-137.
- [6]. Meng L, Feng H. Analysis of characteristics of QQ group hypernetwork[J]. *Application Research of Computers*, 2018, 35(11): 3259-3262.
- [7]. Zhongqun W, Zu W, Yinliang Q, et al. The analysis of product competitiveness based on supernetwork and product online reviews[J]. *Information Studies: Theory & Application*, 2022, 45(4): 113.
- [8]. Gang L, Zhichao B. Research on evolutionary dynamics of knowledge diffusion based on collaboration hypernetwork[J]. *Journal of the China Society for Scientific and Technical Information*, 2017, 36(03): 274-284.
- [9]. Gaojie W, Zhonglin Y, Haixing, et al. Analysis of hypernetwork characteristics in Tang poems and Song lyrics[J]. *Journal of Computer Applications*, 2021, 41(8): 2432.
- [10]. Percy N, Chuzhanova N, Crofts J J. Complexity and robustness in hypernetwork models of metabolism[J]. *Journal of theoretical biology*, 2016, 406: 99-104.
- [11]. Ramadan E, Tarafdar A, Pothan A. A hypergraph model for the yeast protein complex network[C]//18th International Parallel and Distributed Processing Symposium, 2004. *Proceedings. IEEE*, 2004: 189.
- [12]. Pok G, Shon H S, Kim K A, et al. Notice of retraction: Small-World network properties of protein complexes: node centrality and community structure[C]//2011 5th International Conference on Bioinformatics and Biomedical Engineering. *IEEE*, 2011: 1-4.
- [13]. Feng H, Meng L, Jing Z, et al. Analysis and application of the topological properties of protein complex hypernetworks[J]. *Complex Systems and Complexity Science*, 2019, 15(4): 31-38.
- [14]. Lee G, Ko J, Shin K. Hypergraph motifs: concepts, algorithms, and discoveries[J]. *Proceedings of the VLDB Endowment*, 2020, 13(12): 2256-2269.
- [15]. Bretto A. *Hypergraph Theory*[M]. Springer International Publishing, 2013.
- [16]. Milo R, Shen-Orr S, Itzkovitz S, et al. Network motifs: simple building blocks of complex networks[J]. *Science*, 2002, 298(5594): 824-827.
- [17]. Giurgiu M, Reinhard J, Brauner B, et al. CORUM: the comprehensive resource of mammalian protein complexes—2019[J]. *Nucleic Acids Research*, 2019, 47(D1): D559-D563.
- [18]. Barabási A L, Frangos J. *Linked: the new science of networks science of networks*[M]. Basic books, 2002.