A hypergraph structure-based model for the evolution of drug targeted hypernetworks

Libing Bai[a,b,c], Qihang Zhu[a,b,c], Feng Hu[a,b,c,d,*]

[a]School of Computer Science, Qinghai Normal University, Xining 810008, China; [b]Key Laboratory of Tibetan Information Processing and Machine Translation, Xining 810008, China; [c]State Key Laboratory of Intelligent Information Processing and Applications in Tibetan, Xining 810008, China; [d]Academy of Plateau Science and Sustainability, Xining 810016, China

ABSTRACT

Based on the traditional drug-target interactions, a drug-target hypernetwork evolution model was constructed using hypergraph theory. The evolutionary law of the growth of drug-target interactions was analyzed by mean-field theory, and it was found that the distribution of drug-target hypernetwork conformed to a power-law distribution, and further theoretical analysis obtained that the power exponent of the distribution was correlated with the growth rate of the target species corresponding to drug development. A larger exponent tends to explore new targets. By analyzing the drug target data collected from the drugbank in 2021, it was confirmed that the empirical results are consistent with the theoretical analysis and simulation results.

Keywords: Hypergraph, model evolution, empirical analysis

1. INTRODUCTION

With the progress of the times and the rapid development of society, there have been a large number of complex networks in the real world. The emergence of these networks has aroused great enthusiasm among scholars, and the study of complex networks was initiated by the famous mathematicians Erdös and Rényi who proposed the ER random graph[1] model in the 1960s, and the small-world model[2] and scale-free model[3] proposed by Watts and Albert at the end of the 20th century, which started the research of complex networks. Nowadays, the research field of complex networks has been extended to all aspects of social life, such as protein networks, metabolic networks, gene co-expression networks in life sciences, subway networks, airline networks, trade networks in transportation networks[4-7].

The empirical analysis of the evolution patterns of complex network structures and the corresponding modeling studies are the basis for a comprehensive understanding of complex network functions and their applications. For example, the new weighted directed network evolution model proposed by Gao[8] portrays the generation process of new directed edges based on the strength of incoming and outgoing nodes caused by the addition of new nodes, as well as the dynamic changes of the local directed edge weights of the network; You[9] established a technology evolution network model based on patent citation network and used the method of complex networks to reveal the topology, evolutionary rules and dynamics of technology evolution network; Li[10] proposed an evolutionary analysis model of power user groups based on complex networks; Lv[11] proposed a new parallel evolutionary model for complex networks, which allows different localities in the same network to follow different evolutionary models, thus comparing the advantages and disadvantages of different evolutionary models in terms of search efficiency on a unified basis; Cui[12] defined the attractiveness of national LNG trade based on the core drivers (international community's emission reduction policies, supply and demand) in the current international LNG trade, and further constructed a weighted BBV network evolution model based on the attractiveness and distance of national LNG trade through a complex network approach to analyze the evolution of the international LNG trade network from 2016 to 2018, The validity of this evolutionary model is verified.

In the current study of complex networks, the structure of ordinary graphs is commonly used; however, complex networks based on ordinary graphs have certain limitations and duality in describing more complex relationships, while hypergraphs differ from traditional network connections in this mode of connection that conveys higher-order information, and their superiority lies in the fact that the description of hypergraphs is more liberal, because it is not

*qhhuf@163.com

specified that a hyperedge must contain several vertices, so it is relatively ideal for semantic delineation or partitioning of graph data are relatively ideal. For example, in a research collaboration network, assuming that edges are articles and points are article authors, multiple authors of the same article are easily lost in a simple graph. Because a simple graph can only be two points and one article can only connect two authors; however, for hypergraphs, using their properties can better describe the cooperation relationship between authors, so scholars try to extend the traditional complex network theory to the field of hypernetworks. There are two kinds of hypernetworks, network-based supernetwork and hypergraph-based hypernetwork.

As a complex network with multi-level, multi-dimensional and multi-association relationships, hypernetworks have been widely used in research directions such as social network analysis, supply chain, knowledge management and public opinion analysis. Based on the perspective of dependency-based multilayer network, Wang[13] used the patent application data of the Yangtze River Delta (YRD) city cluster for collaborative innovation between industries, universities and research institutes from 2009 to 2018 to construct a multilayer network including cooperation network, knowledge network and city-knowledge element affiliation network, explore the structural characteristics evolution of the collaborative innovation network of the YRD city cluster, and construct an ERGM model to study the formation mechanism of the YRD city cluster collaborative innovation network; Shen[14] considers enterprises as nodes and inter-enterprise relationships as connected edges, and establishes a supply chain financial network evolution model based on the real transaction context of supply chain finance. The cumulative steady-state average degree distribution of the supply chain financial network is obtained through mathematical analysis, and the results are analyzed by computer simulation; Tang[15] constructed a hyper-network dynamic evolution model for the growth and fading mechanism of knowledge in mass collaborative innovation communities; Suo[16] used a hyper-network approach to construct a model based on users' rating data on social networks in order to analyze the characteristics of resource reviews on online social networks and their evolution laws; Liu[17] proposed and established four kinds of small-world model and scale-free model based on a mixture of three-layer hypernetwork evolution model; Hu[18] constructed a research collaboration hypernetwork evolution model based on the collaboration of scientific paper authors using hypergraph theory; Chen[19] constructed a public transportation hypernetwork evolution model based on the evolution characteristics of realistic public transportation system and analyzed the distribution of hyperedges in the evolution and the influence of evolution model parameters on the distribution of hyperedges; Hu[20] constructed a protein complex hypernetwork model and identified key proteins of the network based on the relevant topological indicators of the hypernetwork; Wei[21] introduced the interaction of geographic proximity and technological proximity to form a multidimensional proximity and super-degree combination merit connection mechanism, and constructed a super-network evolution model of emerging technology innovation. The evolution law is revealed through numerical simulation, and the empirical analysis is carried out in the context of China's new energy vehicle technology innovation network.

2. DEFINITION OF HYPERGRAPH

Let $V = \{v_1, v_2, \cdots, v_n\}$ be a finite set, and if $E_i \neq \phi (i = 1, 2, \cdots, e)$ and $\bigcup_{i=1}^{e} E_i = V$, then the binary relation $H = (V, E)$ is said to be a hypergraph. The elements $v_1, v_2, \cdots, v_n$ of V are called the nodes or vertices of the hypergraph and the elements $E_i = \{v_{i_1}, v_{i_2}, \cdots, v_{i_j}\} (1 \leq j \leq n)$ in E are called the hyperedges of the hypergraph. The number of hyperedges containing vertex v_i is called the hypergraph degree of vertex v_i and is denoted as $d_H(v_i)$. The hypergraph H is represented by a graph. In the hypergraph H shown in Figure 1, the vertex set $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$, the hyperedge set $E = \{e_1, e_2, e_3, e_4\}$, where $e_1 = \{v_1, v_2, v_3\}$, $e_2 = \{v_2, v_3\}$, $e_3 = \{v_3, v_5, v_6\}$, $e_4 = \{v_4\}$, and v_7 are isolated vertices, where the supremum of vertices $v_1, v_2, v_3, v_4, v_5, v_6$ are 1, 2, 3, 1, 1, 1 respectively.
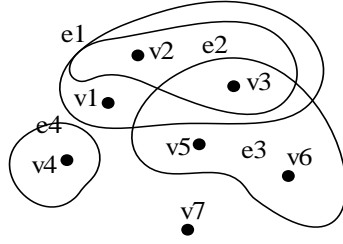
Figure 1. General representation of the hypergraph H.

3. EVOLUTIONARY MODEL OF DRUG TARGET HYPERNETWORK

With the gradual improvement of next-generation high-throughput proteome sequencing technology, the interactions between proteins are becoming more and more perfect, making the identification of drug targets more valuable. Using the interaction network between proteins, the similarity network of drugs, and the targeting network between drugs and proteins, new drug targets can be predicted more accurately; therefore, the construction of drug target hypernetwork can provide new research ideas for predicting drug target modules from a high-dimensional, multi-level, and multi-correlation perspective. Here we give the hypernetwork representation of drug targets, where vertices denote targets and drugs denote hyperedges. For each additional class of drugs, there may be drugs in known hyperedges that act on it, or new target nodes may be added. Different hyperedges are adjacent to each other through common vertices, so a large number of drug target interaction relationships will form a drug target hypernetwork based on drugs with similar potency. Based on this, the construction process of the hyper-network evolution model in this paper is as follows.

(1) Initialization

Assume that initially there are n_0 targets in the network and one drug acting together forms a hyperedge.

(2) Hyperedge growth

A drug is added in each time step according to the variation of time $t(t = 1, 2, \cdots)$ with the following addition rules.

(i) denote with probability P_i the probability that there are i targets in the added drug, where $\sum_{i=1}^{m} P_i = 1$.

(ii) with probability q_j that the target corresponding to the new drug is the same as the j old targets in the original network, where $(j = 0, 1, 2, \cdots, m)$, $\sum_{j=0}^{m} q_j = 1$. Then $j = 0$ means all the targets in the hyperedge are new targets and $j = m$ means all the targets in the hyperedge are old targets.

(iii) Preferential linkage based on hyperdegree

When we select a newly added drug hyperedge, we usually tend to select the target that already has many drugs acting together as the old target in the new hyperedge, so when a drug is added within each time step, the drug is selected to be connected to the old target i in the original network using the hyperdegree priority connection, and the probability $\prod(i)$ of each selected connected node i is equal to the hyper-degree $d_H(i)$ of node i and the ratio of the sum of the hyperdegrees $d_H(j)$ of the already existing nodes j, i.e.,

$$\prod(i) = \frac{d_H(i)}{\sum_j d_H(j)} \tag{1}$$

where the numerator $d_H(i)$ denotes the hyperdegree of node i, i.e., the number of drugs acting on target i, and the denominator denotes the sum of the hyperdegrees of all nodes, i.e., the total number of drugs acting on all targets already

in the network. Figure 2 simulates the evolution of adding to species 5 drugs, the hollow dots indicate the newly added targets and the dotted curves indicate the newly added drugs.
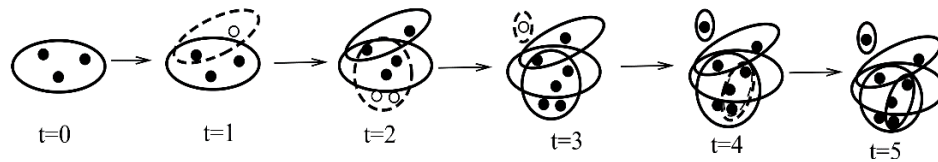

Figure 2. Schematic diagram of the evolution of the drug target hypernetwork.

4. ANALYSIS OF DRUG TARGET INTERACTION DATA

4.1. Data processing and analysis

We constructed the hypernetwork model by processing and analyzing the drug target dataset in drugbank2021.3. The results are shown in Table 1. From Table 1, we know that the types and numbers of drugs acting on multiple targets are much less than those acting on single target, accounting for about 2% of the total, while the number of drugs acting on single target accounts for 76% of the total. The distribution of the number of drugs plotted according to the data in Table 1 is shown in Figure 3, and the cumulative distribution exhibits a clear single-peaked exponential decay function. This is consistent with the analysis of Yildirim et al.[22] in terms of drug targeting relationships, with higher local clustering coefficients indicating a tendency for drugs to connect to older targets. The reason for this trend is that drug studies focus on removing the cause of the disease, restoring the function of the organism, preventing complications, and relieving symptoms. Obviously, prior to clinical studies in humans, all drug effects were based on animal (e.g., rats, dogs, and monkeys) tests, and the effects of drugs on humans and the extent of those effects were speculations and assumptions based on those tests. Therefore, the true effect of a drug on humans must be confirmed by clinical studies. This process is time consuming and heavily regulated, so drugs tend to choose targets that have been experimentally validated.

Table 1. Statistics of the number of targets contained in drugbank-2021.3 drugs.

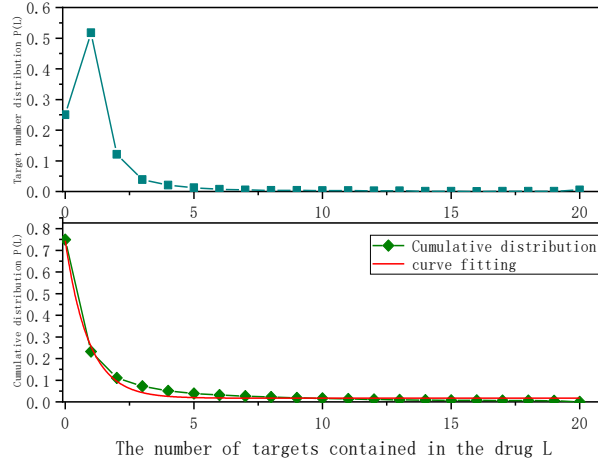| Number of drugs	Drugbank	Proportion (%)
0	1873	25.02
1	3875	51.77
2	908	12.13
3	291	3.89
4	158	2.11
5	93	1.24
6	54	0.72
7	39	0.52
8	28	0.37
9	26	0.35
≥10	140	1.87
Total drugs	7485	

Figure 3. Distribution of the number of targets contained in the drug.

4.2. Mean field theory analysis

In a drug target hypernetwork, we usually focus on how many drugs a target acts with (i.e., the value of the node's hyperdegree), based on which we can judge the drug's the range of drug efficacy, the strength of drug efficacy, etc., and the range of hyperdegree may be related to the tissue-specific region of the disease action. The following is a theoretical analysis of the evolution law of drug target hypernetwork using mean field theory.

The main idea of mean field theory embodied in this paper is: firstly, establish that the hyperdegree $d_H(i)$ of node i satisfies the kinetic equation, secondly, make i obey uniform distribution, reflecting the random selection of node i, then to find out the probability distribution $P_H(d_H, t)$, and finally to take the limit to get the final probability distribution $P_H(d_H)$.

According to the above algorithm for the construction of the hypernetwork model, in each time step t, the addition of a drug (hyperedge) with $l(l=1,2,\cdots,m)$ targets in the drug (determined by the probability p_l) and $j(j=1,2,\cdots,m)$ old targets in the l targets (determined by the probability q_j) causes a change in the hyperedge of node i. The hyperedge of node i, $d_H(t)$ satisfies the kinetic equation.

$$\frac{\partial d_H}{\partial t} = \left(\sum_{l=1}^{m} p_l\left(\sum_{j=0}^{l} q_j j\right)\right)\Pi(i) \tag{2}$$

where $\Pi(i) = \dfrac{d_H(i)}{\sum\limits_{j} d_H(j)}$, and the denominator is the sum of the hyperdegrees of all nodes at moment t, which is the sum of the drugs corresponding to all targets at moment t.

$$\sum_{j} d_H(j) = n_0 + t\left(\sum_{l=1}^{m} p_l l\right) \tag{3}$$

Then

$$\sum_{j=0}^{l} q_j j = J, \quad \sum_{l=1}^{m} p_l J = M, \quad \sum_{l=1}^{m} p_l l = L \tag{4}$$

Equation (2) can be reduced to

$$\frac{\partial d_H}{\partial t} = M\Pi(i) = M\frac{d_H}{n_0 + tL} \tag{5}$$

When t is sufficiently large,

$$\frac{\partial d_H}{\partial t} \approx \frac{M}{L} \cdot \frac{d_H}{t} \tag{6}$$

Each node joins the hypernetwork is the initial value of the node's hyperdegree $d_H(t_i) = 1$, then solving the equation yields $d_H(t) = (\frac{t}{t_j})^{M/L}$, and since the nodes joining the hyperedge in the hypernetwork are chosen randomly, the probability that the node has hyperdegree d_H is:

$$P_H\left(d_H(t) < d_H\right) = P_H\left(t_i > \frac{t}{d_H^{L/M}}\right) \tag{7}$$

Assuming that at the same time interval, the newly added hyperedges obey a uniform distribution, i.e., the value of t_i has a constant probability density $\rho(t_i) = \frac{1}{t}$, thus substituting into the equation to obtain:

$$P_H\left(t_i > \frac{t}{d_H^{L/M}}\right) = 1 - P_H\left(t_i \leqslant \frac{t}{d_H^{L/M}}\right)$$
$$= 1 - \frac{t}{d_H^{L/M} \cdot t} \tag{8}$$

Deriving the above equation, the instantaneous hyperdegree distribution $P_H(d_H, t)$ of the network is:

$$P_H\left(d_H, t\right) = \frac{\partial P\left(d_H(t) < d_H\right)}{\partial d_H} = \frac{L}{M} d_H^{-(1 + L/M)} \tag{9}$$

That is, the node hyperdegree distribution $P_H(d_H) = \frac{L}{M} d_H^{-(1 + L/M)}$.

This shows that the node hyperdegree distribution conforms to the power-law distribution with scale-free property as the degree distribution of many complex network models. The power exponent γ is $1 + \frac{L}{M}$, and a larger γ indicates a faster growth of the drug species acting with that target.

4.3. Monte Carlo numerical simulation analysis

Based on the previous model construction algorithm, a Monte Carlo numerical simulation method was used to generate the number of targets of drug action and the number of connected old targets, and time series were generated iteratively to gradually construct a drug target hypernetwork evolution model, and the distribution characteristics of drug targets over time were simulated by computer.

We suppose there is only one drug at the beginning and the number of targets is 4. With the change of time t, one drug is added in each time step (assuming that the number of targets corresponding to the drug does not exceed 10), and according to the statistics of the number of targets of the drug in Table 1, the probabilities of targets in the drug are taken as $p_1 = 0.25$, $p_2 = 0.52$, $p_3 = 0.12$, $p_4 = 0.04$, $p_5 = 0.02$, $p_6 = 0.01$, $p_7 = 0.01$, $p_8 = 0.01$, $p_9 = 0.01$, $p_{10} = 0.01$. The probabilities of the selected connected old targets were taken as a set of values $q_0 = 0.01$, $q_1 = 0.02$, $q_2 = 0.03$, $q_3 = 0.04$, $q_4 = 0.05$, $q_5 = 0.85$; (we used a computer simulation of the total number of drugs N 7400 in a double

logarithmic coordinate system. The distribution of the number of drugs acting on the target follows a power-law distribution, showing the scale-free property with a power exponent of $\gamma=2.05763$ (Figure 4).
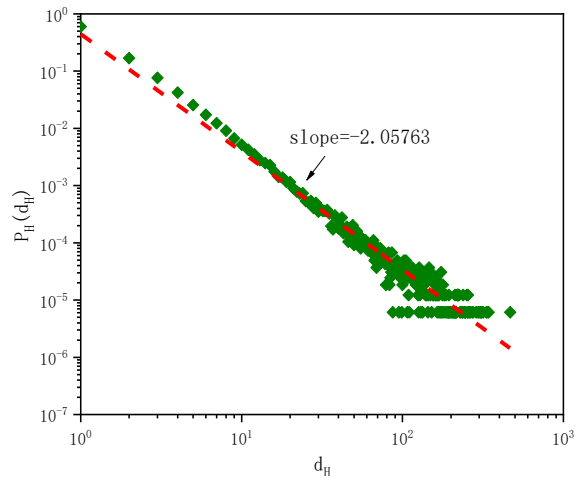

Figure 4. Distribution of hyperdegree of model evolution results with theoretical analysis under specified parameters, simulation results are the average of 50 simulation results.

4.4. Empirical data analysis

To apply the empirical data to validate the evolutionary model in this paper, we collected the drug target dataset of drugbank 2021.3 and analyzed the number of drugs corresponding to the targets, as shown in Table 2.

Based on the data in Table 2, the distribution of the number of target-acting drugs was plotted in a double logarithmic coordinate system as in Figure 5, and the power exponent was found to be consistent with the theoretical and simulated results of the simulation.

Table 2. Statistics of the number of drugs corresponding to drugbank-2021.3 targets.

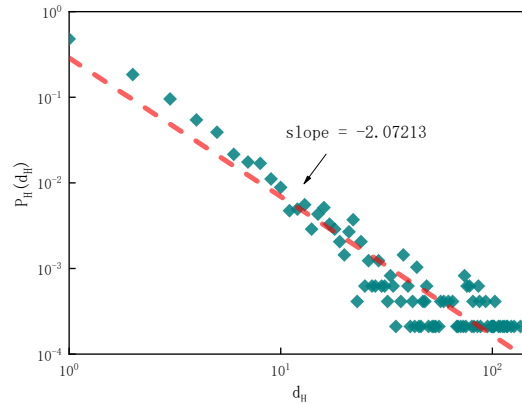| Number of targets	Drugbank	Proportion (%)
1	2018	51.90
2	750	19.29
3	362	9.31
4	203	5.22
5	139	3.58
6	72	1.85
7	54	1.39
8	51	1.31
9	42	1.08
≥ 10	197	5.07
Total targets	3888	

Figure 5. Empirical map of target hyperdegree distribution.

4.5. Comparative analysis of theoretical results

The distribution index $\gamma=1+\dfrac{L}{M}$ of the number of drugs corresponding to the targets (node hyperdegree) is obtained

from the theoretical analysis and defined in Section 4.2: $L=\sum\limits_{l=1}^{m} p_l l$ indicates that there are L targets in a drug added at

time t, $M=\sum\limits_{l=1}^{m} p_l \sum\limits_{j=0}^{l} q_j$, M indicates that there are L targets in a drug added at time t The number of targets in the

existing network is the number of M targets. Obviously, $L \geq M$, the extreme case, when $L=M$, means that the number of target species contained in the added drug is similar compared with the existing ones in the original network, and the index of its target hyperdegree distribution reaches the minimum value $\gamma=2$; when $L>M$, the larger the

proportion occupied by L, the larger the value of γ, i.e., the larger the proportion of old targets in the added drug; on

the contrary, if the number of new targets connected in the added drug during the evolution On the contrary, if the more the number of new targets are connected in the new drugs during the evolution, i.e., each time a new super-edge is

introduced, more and more new nodes are added, the larger γ is. Because $\dfrac{L}{M} \leq L \leq m$ (m is the maximum number of

targets contained in all drugs), then $2 < \gamma \leq 1+m$.

To examine the correlation between drug target growth rate and hyperdegree, this paper compared the number of drug growth with the number of target growth, and the empirical data analysis is shown in Figure 6. Figure 6 shows that the drug target growth rate (i.e., L/M) is 1.05, which translates into a drug target hypernetwork distribution index $\gamma = 1+L/M = 2.05$, consistent with the actual drug target hyperdegree power law index.
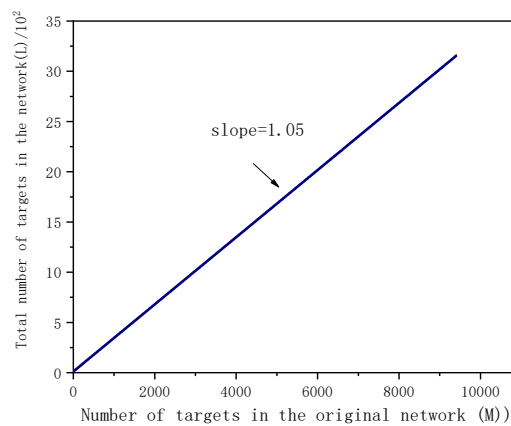

Figure 6. Empirical relationship between targets (M) and the total number of cumulative targets (L) in the original network.

5. CONCLUSION

In this paper, we constructed a drug target hypernetwork evolution model based on drug target interactions according to hypergraph theory, and on this model. We analyzed the evolution law of target drug action by theory, and found that the distribution of target hypernormality follows a power law distribution, and this result is consistent with the results of Monte Carlo simulation experiments and empirical analysis. We also concluded that the value of γ is related to the target growth rate, and the larger γ is, the greater the rate of newly developed drugs corresponding to new targets, and this conclusion is also supported by the empirical data set from drugbank.

ACKNOWLEDGMENTS

REFERENCES

[1] Erdös, P. and Rényi, A., "On the evolution of random graphs," Publ. Math. Inst. Hung. Acad. Sci., 5(1), 17-60 (1960).

[2] Watts, D. J. and Strogatz, S. H., "Collective dynamics of 'small-world' networks," Nature, 393(6684), 440-2 (1998).

[3] Barabási, A. L. and Albert, R., "Emergence of scaling in random networks," Science, 286(5439), 509-512 (1999).

[4] Qiao, S. J., Guo, J., Han, N., Zhang, X. S., Yuan, C. A. and Tang, C. J., "Parallel discovery algorithms for large scale complex network communities," Chinese Journal of Computers, 40, 687-700 (2017). (in Chinese)

[5] Lin, J. X., Hu, Z. S. and Lu, T., "Characterization of protein networks," Journal of Fuzhou University (Natural Science Edition), 486-490 (2014). (in Chinese)

[6] Huang, H. B., Yang, L. M., Wang, J. X., Chen, J. E., Li, S. H. and Du, X. Y., "Research progress on key node identification of biological networks based on network topology," Mathematics in Practice and Theory, (7) 114-125 (2011). (in Chinese)

[7] Wei, Z. B. and Gou, J., "Application and discussion of complex network theory in grid analysis," Power System Technology, 59, 279-287 (2015). (in Chinese)

[8] Gao, Q. Y. and Li, M., "A new evolution model for weighted directed networks model," Journal of Northwestern Polytechnical University, 38(4) 913-917 (2020). (in Chinese)

[9] You, G., Guo, H. and Liu, X., "Technology evolution network model and simulation based on patent citation network," Journal of System Simulation, 33(3), 591-603 (2021). (in Chinese)

[10] Li, C. X., Shi, J. Q., Liu, N., Ma, L. Y. and Li, C. C. "An evolutionary analysis model of power user groups based on complex networks," Chinese Journal of Electrical Engineering, 7, 1-14 (2021). (in Chinese)

[11] Lv, T. Y., Huang, S. B., Park, X. F., Gao, K. and Jia, Y. R., "Parallel evolutionary analysis of complex networks with multiple models based on search efficiency," Chinese Science: Physics Mechanics Astronomy, 21(2), 159-166 (2013). (in Chinese)

[12] Cui, W., Xia, R., Yang, L. L. and Sun, J. Q., "Research on the evolution of complex network of international LNG trade," Journal of Shenyang Agricultural University, 51, 394-401 (2020). (in Chinese)

[13] Wang, H. H., Sun, Q., Du, M. and Li, Y., "Research on the evolution and formation mechanism of collaborative innovation network in Yangtze River Delta city cluster—A perspective of dependent multilayer network," Science & Technology Progress and Policy, 37(9), 69-78 (2020). (in Chinese)

[14] Shen, A. Z., Guo, J. L., Suo, Q. and Wang, F. H., "Modeling and analysis of supply chain finance based on multilayer network," Application Research of Computers 34(12) 3628-3631 (2017). (in Chinese)

[15] Tang, H. T. and Li, C. H., "Community knowledge discovery and analysis based on a hypernetwork evolutionary model," Systems Engineering—Theory & Practice 38(3) 765-776 (2018). (in Chinese)

[16] Suo, Q. and Guo, J. L., "A hypernetwork evolutionary model of online social network resource review relationship," Journal of Systems & Management 25(5) 852-857+867 (2016). (in Chinese)

[17] Liu, Q., Fang, J. Q. and Li, Y., "Characterization of a three-layer hypernetwork evolutionary model," Complex Systems and Complexity Science, 12(2), 64-71 (2015). (in Chinese)

[18] Hu, F., Zhao, H. X., He, J. B., Li, F. X., Li, S. L. and Zhang, Z. K., "Evolutionary model of research collaboration network based on hypergraph structure," Acta Physica Sinica, (19), 547-554 (2013). (in Chinese)

[19] Lu, R. M. and Guo, J. L., "Construction and analysis of evolutionary model of urban bus hypernetwork," Software Guide, 164-166+172 (2019). (in Chinese)

[20] Hu, F., Liu, M., Zhao, J. and Lei, L., "Analysis and application of protein complex hyper-network properties," Complex Systems and Complexity Science, 15, 31-38 (2018). (in Chinese)

[21] Wei, X. G., Luo, Y. K., Yang, Q. S. and Zhang, Z. Y., "A study on the evolution of emerging technology innovation hyper-network based on multidimensional proximity—Taking the new energy automobile industry as an example," Industrial Technology Economics, 56-64 (2021). (in Chinese)

[22] Yildirim, M. A., Goh, K. I., Cusick, M. E., Barabási, A. L. and Vidal, M., "Drug-target network," Nature Biotechnology, 25(10), 1119-1126 (2007).