



Heuristic Network Similarity Measurement Model Based on Cloud Computing

Yang Guo^{1(✉)} and Jia Xu^{1,2}

¹ Dalian Institute of Science and Technology, Dalian, China
yc306@le.ac.uk, jf125125@163.com

² Dalian Jiaotong University, Dalian, China

Abstract. In order to solve this problem, a heuristic network similarity measurement model based on cloud computing is proposed. First, the heuristic network data is collected, and then the spherical harmonic function method is used to match the network data similarity measurement. After the above work, the heuristic network similarity measurement model is built according to the structure balance theory. Thus, a heuristic network similarity measurement model based on cloud computing is constructed. In the experiment, the quality of service node genes obtained by the two models was tested. The experimental results show that the service node genes obtained by the model are better and meet the design requirements.

Keywords: Cloud computing · Heuristic network · Similarity measurement · Node · Evaluation criteria · Network data · Similarity propagation mode

1 Introduction

At present, the research on polar information network mainly focuses on link prediction. Among them, part of the work involves the correlation measurement of nodes, but only uses the correlation measurement method on the non-polar information network, there is no special research on the correlation measurement of nodes on the polar information network by using the negative edge information effectively. Based on the status theory, foreign scholars calculate the similarity between two nodes according to the positive degree, negative degree, positive degree and negative degree. Positive in and negative in can improve the status of nodes, while positive in and negative in can reduce the status of nodes. However, only the degree of two nodes is considered, and whether there is a relative preference relationship between two nodes is not considered [1].

In this paper, we focus on the measurement of the correlation between nodes in polar information networks, including the measurement of the similarity between nodes of the same type in homogeneous polar information networks and the measurement of the correlation between nodes of different types in heterogeneous polar information networks. The main work and contributions of this paper are summarized as follows: in order to test the similarity between nodes of the same type in homogeneous polar information networks, a heuristic network similarity measurement model based on

cloud computing is proposed. The model can make full use of the semantic information contained in the positive and negative sides of the isomorphic polar information network. The similarity between the two nodes is measured by comparing the direct neighbor sets of the source node and the target node. Because there are both positive and negative edges in polar information networks, we need to consider the positive and negative neighbor sets of nodes respectively. Furthermore, for the directed isomorphic polar information network, the neighbor set is subdivided into positive incoming neighbor set, negative incoming neighbor set, positive outgoing neighbor set, and negative outgoing neighbor set due to the more definite direction of the edge. For two nodes without common neighbors, similarity propagation is used to measure the similarity between the source node and the target node. Finally, experiments are carried out on real datasets, and the experimental results show the effectiveness of the proposed method.

2 Heuristic Network Data Collection

Heuristic network data collection refers to the process of constructing a heuristic network data with some algorithm for the original feature set of input data set. The heuristic network data extracted from the original feature variables is the most consistent with the set feature selection criteria [2]. The data with lighter weight in the original data can be filtered by feature selection, and the data that can best reflect the data features will be preserved. After feature selection, the data model will be more accurate and simple, and the processing efficiency of the model will be greatly improved. Feature selection model can be described as a simple mathematical model: Given the sample data set $S = \{F, C, D\}$, where F represents the feature sample set, C represents the category sample set, D is the data sample set. It is assumed that the feature algorithm $E(x) \in (0, 1)$ is a feature evaluation function, and its value size corresponds to the weight of data, which measures the importance of data. There are several types of functions in the process of selecting optimal heuristic network data:

- A: Select a subset in feature set F to make $E(x)$ maximum;
- B: The minimum value $E(x)$ of the given E_0 value. The corresponding function value is greater than the subset.
- C: Find a subset in the feature set F so that the amount of $E(x)$ is large and the number of features is as small as possible;

The above methods take different measures from the point of view of the number or weight of features, but ultimately can select the most favorable feature data [3].

Heuristic network data collection process (Fig. 1);

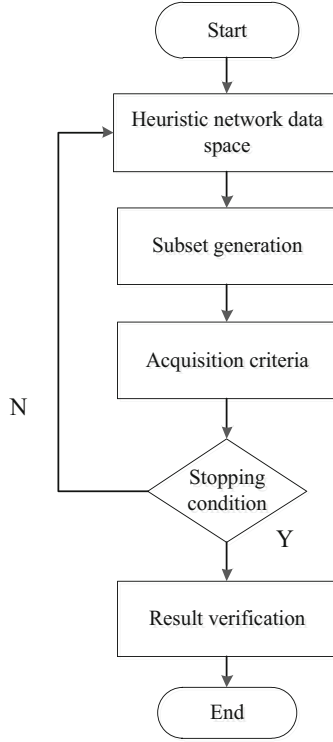


Fig. 1. Heuristic network data collection process

The subset generation part refers to the process of extracting a subset from the original feature set, which generally needs to search for a subset that can meet specific conditions in the feature space according to some search rules. This section contains two key concepts: search direction and search strategy. The search direction refers to the starting point of feature search and the direction of search [4]. Any subset can be used as the starting point of search, which determines the search direction: forward search (starting from the empty set), backward search (starting from the whole feature set), two-way search (starting from both directions above), and random search (starting from a randomly generated subset). Search strategies generally include complete search, randomized search and heuristic search [5]. Evaluation criteria refer to the criteria for evaluating the heuristic network data selected from the generated part, mainly focusing on the rationality of the results, the effectiveness of the algorithm in the problem and whether it is helpful to achieve the specified goals. Different evaluation criteria will lead to different heuristic network data, and the quality of evaluation criteria will directly affect the effect of the algorithm. Common evaluation criteria can be divided into distance measurement, information measurement, consistency measurement, classification error rate, etc. For different specific problems, different

evaluation criteria can be selected [6]. Stop condition can also judge the rationality of current heuristic network data and send the unreasonable heuristic network data back to the subset generation process for screening again. Generally, the stop condition can be set as follows:

- A: The number of subsets reaches the preset value.
- B: According to the currently selected heuristic network data, the classification rate has reached the requirements or been improved.
- C: The change of feature number no longer affects the function value of evaluation criteria.

The value of the D: evaluation criterion function reaches the inflection point or the threshold [7] already set.

- E: The selected heuristic network data is the optimal solution of the evaluation criteria function.

The result verification part refers to comparing the optimal heuristic network data selected from the original features and the existing optimal subset, which has been used to verify the applicability and rationality of the algorithm, so that the optimal algorithm and the corresponding optimal heuristic network data can be determined.

3 Cloud Computing Heuristic Network Similarity Measurement Matching

Firstly, the heuristic network data is collected, and then to solve the problem of spherical harmonic function information loss, the cloud computing heuristic network similarity measurement matching is realized. At present, the proposed solutions can be roughly divided into three categories:

- A: Some changes are added when using the spherical harmonic function method to retain some local information, such as the optical ray method and the improved optical ray method.
- B: When Chebyshev points are extracted by spherical harmonic decomposition, the points containing global information are sampled.
- C: Map the decomposed harmonic components to the Cartesian coordinate system, and add a coordinate [8].

These three methods have their own advantages and disadvantages: the first method has high execution efficiency, but it can not eliminate the problem of information loss in essence; the second method is novel and effective, but unstable; the third method is the best method among the current three-dimensional object shape matching methods, but the algorithm complexity is high, which is not suitable for large database query [9]. Because of the high information stored in 3D object model, the efficiency is very important in 3D object query system. Based on this consideration, this paper decided to

adopt the first method and improve the matching effect as much as possible on the basis of high efficiency.

Cloud computing heuristic network similarity measurement matching process (Fig. 2):

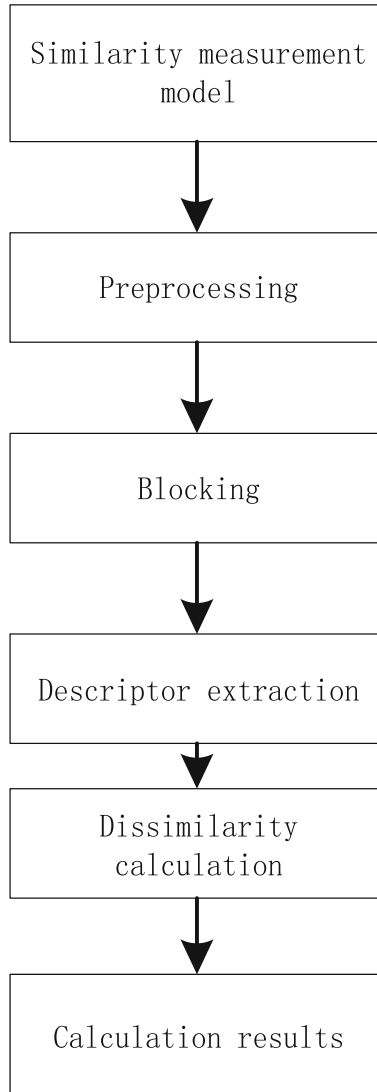


Fig. 2. Cloud computing heuristic network similarity measurement matching process

In the first step, the preprocessing process first needs to express the ball function in the form of $f(\theta, \varphi) = R$ where R is the distance from the sample point of Monte Carlo surface to the center of mass of the object [10].

Secondly, in order to find the corresponding sector blocks of two objects, the spherical function $f(\theta, \varphi) = R$ should be normalized by PCA coordinate axis transformation.

In the third step, the 3D model needs to be divided into n sector blocks, as shown in Fig. 3.

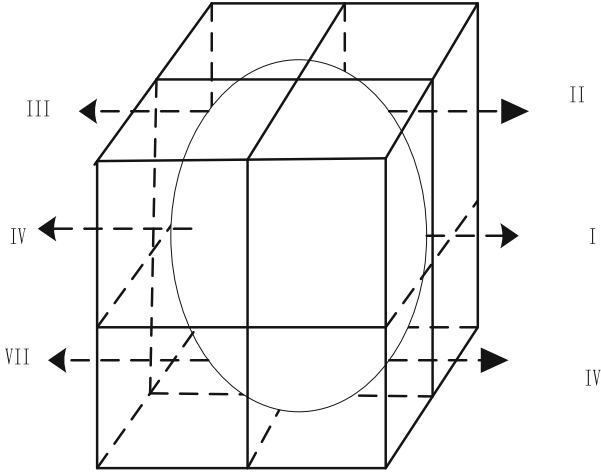


Fig. 3. Block by block diagram

First, the ball function is divided into eight parts, that is, eight quadrants, Then divide each quadrant into $\frac{n}{8}$ blocks by dividing φ equally.

As the block is carried out in the new coordinates, it is necessary to transform the coordinates of each point. In this paper, the affine coordinate transformation method is used, and the affine coordinate system is a special case of rectangular coordinate system. After the original spherical function $f(\theta, \varphi) = R$ is partitioned according to the coordinate value of the point after coordinate transformation, the points of the spherical function are stored in n arrays for further processing. In the fourth step, the spherical harmonic function of N sector blocks of the spherical function is decomposed. After spherical harmonic function decomposition of each sector block, the obtained rotation invariant descriptor is:

$$SH(f)_n = \{ \|f_0(\theta, \varphi)\| \} \tag{1}$$

In practical operation, it is impossible for us to obtain the energy component with infinite frequency. As an approximation, we can decompose the spherical coordinate function into the sum of the finite terms of the spherical harmonic function with a certain bandwidth, B is used to specify the upper limit of the frequency, that is, the bandwidth, and the higher expansion coefficient exceeding the frequency will be ignored:

$$f_0(\theta, \varphi) = \sum_{l=0}^{B-1} a_{l,m} \quad (2)$$

Through a large number of tests, it is found that when $B = 64$ or greater, the accuracy of the system has generally reached about 5×10^3 . For the requirements of 3D model retrieval, it is enough.

Finally, by summarizing the descriptors of these n sectors, we get a two-dimensional grid distribution, which is the final shape descriptor.

Fourth, there are many ways to match the final descriptor. A simple method is to directly subtract the descriptors of the corresponding sector block, and then overlay the results of the subtraction to get the final dissimilarity.

4 Heuristic Network Similarity Measurement Model

First, the heuristic network similarity data is collected, then the heuristic network similarity measurement is matched, and the heuristic network similarity measurement model is constructed by using the structure balance theory.

Specific implementation process:

In an undirected homogeneous polar information network, given two nodes v_i and v_j , by considering their direct neighbor set, we give the calculation method of measuring the similarity between these two nodes. For node VI on undirected homogeneous polar information network, its direct neighbor set, the direct neighbor set connected with VI positive edge and the direct neighbor set connected with VI negative edge are respectively:

$$\begin{aligned} N_i &= \{v_k | (v_i, v_k) \in \delta\} \\ N_i^+ &= \{v_k | (v_i, v_k) \in \delta^+ \} + \{v_i\} \\ N_i^- &= \{v_k | (v_i, v_k) \in \delta^- \} \end{aligned} \quad (3)$$

Among them, $N_i = N_i^+ \cup N_i^-$.

To sum up, the more common neighbors with the same evaluation between two nodes, the more similar they are. Accordingly, the more common neighbors with

opposite evaluation between two nodes, the more dissimilar they are, given two nodes v_i and v_j . We can get the set of neighbors with the same evaluation and the set of neighbors with different evaluation.

$$C_S(v_i, v_k) = \left\{ v_k \mid \left(v_k \in N_i^+ \wedge v_k \in N_j^+ \right) \right\} \quad (4)$$

Where $C_S(v_i, v_k)$ represents the set of common neighbor nodes with the same evaluation held by v_k and Represents a set of common neighbor nodes with different evaluations held by v_k and v_i .

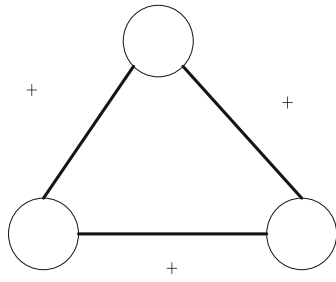
Then the similarity between the two nodes is defined as follows:

$$sim(v_i, v_k) = \frac{|C_S(v_i, v_k)| - C_d(v_i, v_k)}{|N_i \cup N_j|} \quad (5)$$

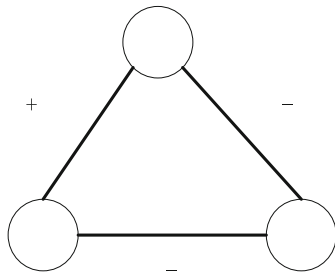
The value range of the similarity between the two nodes is in the range of $[-1, 1]$, $sim(v_i, v_k) > 0$ Indicates that nodes v_k and v_i are similar, $sim(v_i, v_k) < 0$ Indicates that the nodes v_k and v_i are not similar. The greater the similarity value is, the more similar the two nodes are, and the smaller the value is, the less similar the two nodes are. It should be noted that both the maximum value 1 and the minimum value -1 can be obtained. When two nodes have identical neighbors, and their evaluation of each neighbor is the same, the similarity between the two nodes is 1. When two nodes have identical neighbors, but their evaluation of each neighbor is not the same, then the similarity between the two nodes is -1 , similarity 1 means that the positions of the two nodes are identical, and similarity -1 means that the positions of the two nodes are different.

The similarity measure is symmetric and satisfies $sim(v_i, v_k) = 1$, It is worth noting that if we ignore the information, we define the similarity on the undirected isomorphic polar information network, and the measurement method is the same as Jaccard coefficient.

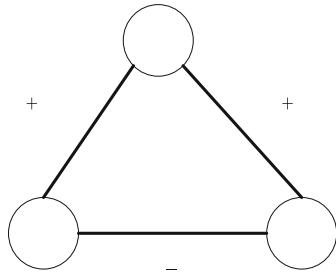
Use formula (3) to calculate the similarity of nodes in Fig. 4, and consider users $u1$ and users $u3$. Then $N_1 = \{U1, U2, U3, U4, U6, U7\}$. The similarity between user $U1$ and user $U2$ is $sim(U1, U2) = -\frac{1}{3}$, According to the theory of structural equilibrium, it can be seen in Fig. 4.



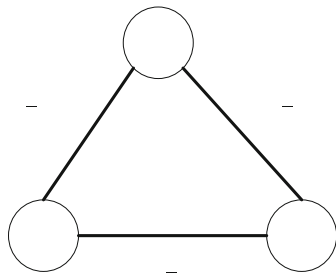
(a)



(b)



(c)



(d)

Fig. 4. Structural equilibrium theory

In the directed polar information network, the similarity between two nodes is measured. The directivity of the edge is more clear than that of the undirected information network. A user can express his attitude to other users, and can also accept the evaluation of other users. Therefore, we divide the user's neighbor set into two parts: (1) in neighbor set; (2) out neighbor set. In particular, the incoming neighbor set includes the positive incoming neighbor set and the negative incoming neighbor set, and the outgoing neighbor set includes the positive outgoing neighbor set and the negative outgoing neighbor set. As shown in Fig. 5;

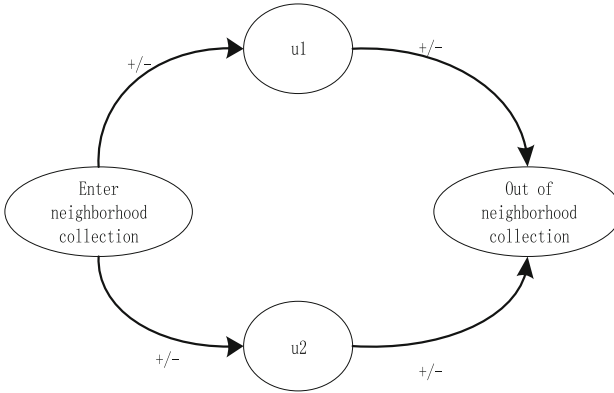


Fig. 5. Heuristic network similarity measurement model

If two users want to have a high degree of similarity, they should meet two conditions: the first condition is that as many users hold the same evaluation as possible; the second condition is that as few users hold different evaluation as possible. Then we can extend this basic idea to apply to the directed polar information network.

If two users want to have high similarity, they need to meet the following conditions:

- A: We should try to evaluate as many users as possible.
- B: We should try to evaluate as many users as possible.
- C: The opposite users should be evaluated as little as possible.
- D: Users with opposite evaluation should be as few as possible.

In other words, if two users have the same evaluation on many other people, and many people have the same evaluation on these two users, then the two users are likely to be highly similar. Correspondingly, if two users have different comments on many other people, and many people have different comments on these two users, then the two users may be different.

Thus, the heuristic network similarity measurement model based on cloud computing is completed.

5 Experimental Analysis

In order to verify the feasibility and effectiveness of the model, this chapter will carry out simulation experiments. The experimental environment is: the experimental host is configured with Intel Core2 2.93 GHz processor, 2.00 gb memory, Windows XP operating system, and the algorithm is implemented with MATLAB 9.

5.1 Experimental Parameters

The heuristic optimization algorithm in this paper consists of two stages. The following describes the setting of key parameters in the two stages:

The main parameters of the traditional model are population size n , crossover probability p_c , mutation probability p_m , where p_c and p_m represent the adjusted probability of white adaptation. The population size is $n = 40$. To facilitate calculation, the size of the element candidate service set C_i in the set C of the candidate service set is the same.

The heuristic network similarity measurement model based on cloud computing influences the behavior of artificial ants by changing the three parameters of information heuristic factor α , expectation heuristic factor β , and pheromone volatility coefficient ρ , we reduce the influence of pheromone on the movement of artificial ants by reducing the value of α and ρ and increase the exploration ability of artificial ants by increasing the value of β , $\beta = 5$.

5.2 Experimental Result

The business process contains four abstract services, and each candidate collection of abstract services is assumed to have the same scale, with five entity Web services respectively. In this paper, we consider two QoS parameters including overhead and response time, and define the global QoS constraint vector as $*R = (400, 90)$. First of all, in the preprocessing candidate service set of similarity measurement parameters, the list of entity service similarity measurement parameters is shown in Table 1. Because the cost and response time belong to the linear additive, negative heuristic network similarity measurement model, there is no need for type conversion, only a unified numerical order of magnitude, mapping to the same range of values. According to the preprocessing operation of application formula (5), the parameter list of entity service heuristic network similarity measurement model is shown in Table 1. In the calculation process of using formula (5), set the real number $C1 = 0$, $C2 = 100$ (Table 2).

Table 1. Entity service heuristic network similarity measurement parameter list

C1		C2		C3		C4	
Expenses	Response time	Expenses	Response time	Expenses	Response time	Expenses	Response time
204	5	10.5	10	33	42	94	20
203	5	10.3	5	43	43	95	30
205	4	19.2	7	51	51	84	13
206	2	14.5	8	46	34	75	21
212	3	6.2	5	42	41	123	16

Table 2. Application formula (5) list of heuristic network similarity measurement parameters after preprocessing

C1		C2		C3		C4	
Expenses	Response time	Expenses	Response time	Expenses	Response time	Expenses	Response time
52	50	15	45	40	65	54	50
56	30	80	0	65	16	100	60
100	0	100	15	46	0	22	45
80	50	60	100	30	162	84	0
5	60	0	90	60	54	0	152

According to the algorithm of initial distribution of pheromones, a group of feasible solutions are obtained, including $(cs_{11}, cs_{22}, cs_{32}, cs_{43})$ and $(cs_{11}, cs_{22}, cs_{32}, cs_{43})$, as shown in Table 3. The aggregation cost of heuristic network similarity measurement parameters is also listed in the table.

Table 3. List of feasible solutions of two models

Feasible solution	Traditional heuristic network similarity measurement model	Heuristic network similarity measurement model based on Cloud Computing
$(cs_{14}, cs_{22}, cs_{32}, cs_{43})$	70	302
$(cs_{11}, cs_{22}, cs_{32}, cs_{43})$	80	352

The comparison results show that feasible solution $(cs_{11}, cs_{22}, cs_{32}, cs_{43})$ is a better execution plan than feasible solution $(cs_{14}, cs_{22}, cs_{32}, cs_{43})$. We set pheromones according to this set of execution plans. At the beginning stage, the initial value of pheromones of all nodes is set to 1.00, and the two feasible solution parts are overlapped, including entity services $cs_{22}, cs_{32}, cs_{43}$. It is clear that the genes of these nodes are better, and the pheromone value of these service nodes should be slightly higher, set to 1.20, cs_{11} as the starting point of the better solution, set cs_{11} , the pheromone value of the service node where is 1.15, and set the sub optimal solution starting service cs_{14} . The pheromone value of the service node is 1.10. The initial distribution of pheromones is obtained by calculation, as shown in Fig. 6.

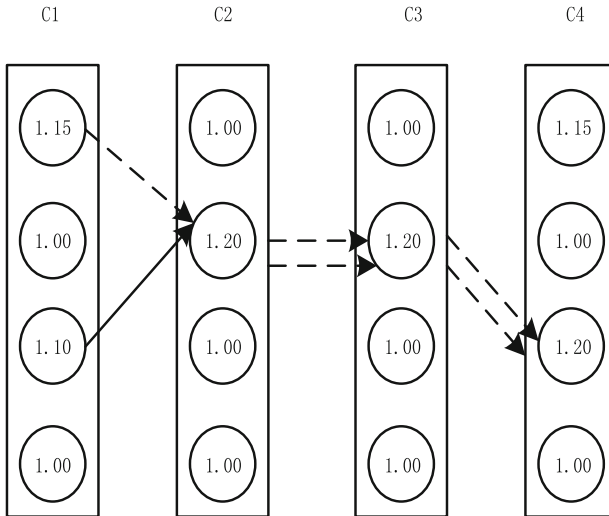


Fig. 6. Initial distribution of pheromones in the model

As shown in Fig. 6, using the heuristic network similarity measurement model of cloud computing, these node genes are better.

6 Conclusion

This paper studies the similarity measurement problem on heuristic networks and discusses it. This paper proposes a new method to measure the similarity between nodes of the same type on the same polar information network. This method can take into account the interaction of positive and negative sides on the polar information network at the same time, and make full use of the semantic information of the polar information network. The experimental results on real datasets also show the effectiveness of the proposed method.

References

1. Chen, X.: Research on vine type serial-parallel clustering method of network public opinion based on cloud computing. *J. Intell.* **36**(2), 122–126 (2017)
2. Li, H., Mao, Y.: Forecasting model of aquaculture production based on heuristic Johnson algorithm optimization and BP neural network. *Fishery Modernization* **44**(6), 19–23 (2017)
3. Yuan, J.M.: Research of collection model based on heuristic algorithm in MANET. *Control Eng. China* **24**(6), 1248–1253 (2017)
4. Song, J., Wen, L.: A similarity measure for process models based on task occurrence relations. *J. Comput. Res. Dev.* **54**(4), 832–843 (2017)
5. Han, X., Wang, Q.: Geographic ontology concept semantic similarity measure model based on BP neural network optimized by PSO. *Comput. Eng. Appl.* **53**(8), 32–37 (2017)

6. Huang, Y., Xie, F.: k-modes algorithm based on structural similarity. *Comput. Eng. Appl.* **53**(23), 102–107 (2017)
7. Yang, S., Liu, C.: Two stages of heuristic based algorithm for influence maximization in social network. *J. Chin. Comput. Syst.* **38**(10), 2268–2274 (2017)
8. 胡启志, 颜娜.: Influence maximization algorithm based on term frequency and node similarity in social networks. *J. Chin. Comput. Syst.* **38**(2), 259–263 (2017)
9. Ye, J.: Clustering algorithms using dimension root distances-based cosine similarity measures of single-valued and interval neutrosophic sets. *Mach. Tool Hydraulics* **46**(6), 199–208 (2018)
10. Yao, Z., Wu, L.: Network visual analysis method based on heuristic community detection. *J. Comput. Appl.* **37**(s1), 155–159 (2017)