Scientific Research

Scientific Communities Found Based on the Path Structure of Citation Network

Xiao Xiao[1,2], Song Cao[1,2], Lan Huang[1,2*], Yutian Tang[1,2]

[1]College of Computer Science and Technology, Jilin University, Changchun, China
[2]Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education,
Jilin University, Changchun, China
Email: [*]huanglan@jlu.edu.cn

ABSTRACT

Based on the structure of citation network, the citation paths among papers, and the association strength such as coupling, co-citation and etc. between two papers are defined in this article. We give formulas to quantify the association strength in order to establish citation network model based on the citation path structure. Then, the OPTICS algorithm is brought into the scientific communities found model since it can solve the parameter's setting problem. This method combines various kinds of path structures together and thus it contains more complete citation network information. Experiments and analysis reveal the reliability and validity of this method.

Keywords: Citation Network; Scientific Communities Found; OPTICS Algorithm; Citation Path

1. Introduction

Many researchers have been studying and mining scientific communities using citation network and analysis of the citation network has also become an important tool for their research. Xiaoming Liu employed the analysis method of social network into the co-author network to detect condition variations of the digital libraries and constructed weighted directed citation network to mine the inner scientific communities [1]. R. Ichise designed a user-interacted community mining system, in which the values of co-author network parameter can be constructed variably in different needs of users' purposes. On that basis, they added the author-keyword schema and make a good attempt at clarifying research interests of different scientific communities [2].

A large amount of previous research has been done on citation analysis method, however, these studies and experiments are all one-sided, doing survey on certain aspects of the research, instead of a comprehensive vision. Shortcomings are listed as follows: First of all, the comparison criterion between relational model is inconsistent. Although the citation path provides association relationship between papers, there's no uniform criterion to quantify this relationship. Secondly, there's no documental multiple correlation. The majority of previous scientific community found are doing survey in a certain aspect of citation path, co-author, coupling, co-citation, and thus it is incomprehensive. Thirdly, the results of the mining groups are impracticality.

This article defines the association strength between authors and the association strength between papers. The formulas of those definitions innovatively combine coupling, co-citation, co-author and other structure to make the system more comprehensive. Besides, the proposed OPTICS algorithm is not sensitive to parameter setting, and it can use heuristic algorithm to estimate the parameter and get good mining results of a large dataset.

2. Build of Citation Network Based on the Citation Path Structure

Three common types of citation network structures are coupling, co-citation, co-author. This article uses weighted directed graphic method to build the network model of citation network. Here are some hypotheses: Firstly, assuming there's no fault reference or repeated reference issues between papers. Secondly, association relations between papers can be measured numerically and the tightness of association relations can also be compared by numeric size.

In citation network, a paper is always regarded as a vertex and a relation as an edge, and thus a mesh structure is formed. Then a citation network model based on the citation path can be built according to the below definitions. The first four definitions below can be found in reference [3].

Definition 1) Citation Network
Definition 2) DR (Direct Reference)
Definition 3) RP (Reference Path)

Definition 4) RL (Reference Length)

Definition 5) RPR (Reference Path Relation)

RPR is described by reference length between two papers.

$$RPR\left(P_i, P_j\right) = \sum_{\mu=1}^{n} \frac{1}{PRL_{\mu}\left(P_i, P_j\right)} \quad (1)$$

where there are n citation paths from paper P_i to paper P_j, and the length of each path is $RL_{\mu}\left(P_i, P_j\right)$. The longer RL is, the less RPR is. The more RP is, the more RPR is. Thus, we use cumulative method.

Definition 6) BCR (Bibliographic Coupling Relation)

BCR is described by the reference path from the coupling relation papers to the co-citation paper.

$$BCR\left(P_i, P_j\right) = \sum_{\mu=1}^{n} \frac{1}{RL\left(P_i, P_{\mu}\right) + RL\left(P_j, P_{\mu}\right)} \quad (2)$$

where n is the number of co-citation papers by P_i and P_j, P_{μ} represents the μ-th paper they co-citation, $RL\left(P_i, P_{\mu}\right)$, $RL\left(P_j, P_{\mu}\right)$ is separately used to represent the reference length form P_i and P_j to P_{μ}. As is shown in the **Figure 1**,

$$BCR\left(P_B, P_C\right) = 1/2 , \quad BCR\left(P_B, P_D\right) = 1/3 .$$

P_D refers P_A indirectly, so the coupling association strength between P_D and P_B is less tight than that of P_C and P_B.

Definition 7) CCR (Co-citation Relation)

CCR is described by the reference paths from the co-citation paper to the two papers respectively.

$$CCR\left(P_i, P_j\right) = \sum_{\mu=1}^{n} \frac{1}{RL\left(P_{\mu}, P_i\right) + RL\left(P_{\mu}, P_j\right)} \quad (3)$$

where n is the number of co-citation papers of P_i and P_j, P_{μ} represents the μth paper they co-citation, $RL\left(P_i, P_{\mu}\right)$, $RL\left(P_j, P_{\mu}\right)$ is separately used to represent the reference length form P_i and P_j to P_{μ}.

Definition 8) PR (Paper Relation)

$$PR\left(P_i, P_j\right) = \lambda_1 RPR + \lambda_2 BCR + \lambda_3 CCR \quad (4)$$

PR is accumulated by weighted RPR, BCR and CCR. These three association strengths present the association relations between two papers from different aspects, thus, their weighted value λ_1, λ_2, λ_3 should be different. In order to discuss the issue, here we set $\lambda_1 + \lambda_2 + \lambda_3 = 1.0$.
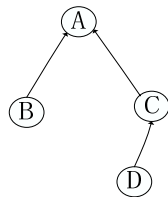

Figure 1. Bibliographic coupling relation.

Definition 9) TR (Together write Relation)

If two authors write a paper together, namely co-author, it demonstrates that they share some interrelated research direction. TR is used to represent this kind of association strength.

$$TR\left(a_i, a_j\right) = \sum_{t=1}^{n} \frac{2}{M_t} \quad (5)$$

where n is the number of paper co-authored by a_i and a_j, M_t represents the number of the authors who co-author the t-th paper, the PR of each paper is $2/M_t$, the overall TR is accumulated by TR of each co-authored paper.

Definition 10) AR (Author Relation)

$$AR\left(a_i, a_j\right) = \beta_1 \sum_{\mu=0}^{m} \sum_{\nu=0}^{n} \frac{PR\left(P_{\mu}, P_{\nu}\right) + PR\left(P_{\nu}, P_{\mu}\right)}{\left|Year\left(P_{\mu}\right) - Year\left(P_{\nu}\right)\right|} \quad (6)$$
$$+ \beta_2 TR\left(a_i, a_j\right)$$

where m is the total number of the papers author a_i has published, n is the total number of the papers author a_j has published, P_{μ} is the μth paper author a_i has published, P_{ν} is the v-th paper the author a_j has published, $Year\left(P_{\mu}\right)$ and $Year\left(P_{\nu}\right)$ separately represent the year of the published paper P_{μ} and P_{ν}. If $Year\left(P_{\mu}\right) = Year\left(P_{\nu}\right)$, then we set

$$\left|Year\left(P_{\mu}\right) = Year\left(P_{\nu}\right)\right| = 1 .$$

If P_{μ} and P_{ν} is the same paper, in other words, it is co-authored, then we set

$$\frac{PR\left(P_{\mu}, P_{\nu}\right) + PR\left(P_{\nu}, P_{\mu}\right)}{\left|Year\left(P_{\mu}\right) - Year\left(P_{\nu}\right)\right|} = 0 .$$

we can use $TR\left(a_i, a_j\right)$ to represent their association strength. If P_{μ} and P_{ν} is not the same paper, then $TR\left(a_i, a_j\right) = 0$, we can also use $TR\left(a_i, a_j\right)$ to represent their association strength. Considering the two parts represent the different association strength, we accumulate them by weights. In order to discuss this issue, here we set $\beta_1 + \beta_2 = 1.0$.

3. Scientific Communities Found Method Based on Citation Path

3.1. OPTICS Clustering Algorithm

DBSCAN algorithm is sensitive to the parameters of ε and Minpts [4]. Even a slight different setting may lead to a different clustering result. Setting of the two parameters requires experience. As for the high-dimensional data, setting appropriate parameters is challenging. In order to solve the shortcomings of DBSCAN algorithm, this article brings in OPTICS [5] (Ordering Points to Identify the Clustering Structure) clustering algorithm [4]. OPTICS algorithm is based on the fact that low-density

clusterings contain high-density clusterings, as shown in the **Figure 2** where ε_1 is the distance parameter of clustering C, ε_2 is the distance parameter from cluster-ing C1 to C2, and $\varepsilon_1 > \varepsilon_2$, clustering C includes cluster-ing C1 and clustering C2 completely.

OPTICS algorithm is the extension of the DBSCAN algorithm [5]. Unlike DBSCAN algorithm which generates clustering result after distributing the objects into different clusters [6], OPTICS algorithm stores a cluster order when each object is being clustered [7].Through the order we can get some basic clustering information such as the center of the cluster and inner cluster structure in the object set of the database. As for each object, these information is made up of two parts: core distance and reachable distance.

Definition 11) Core Distance (CD)

In the object se of the database O, the core distance of object p is defined as: if object p is the core object, $Card\left(N_\varepsilon\left(p\right)\right) \geq Minpts$, then the distance from object p to its Minpts nearest surrounding object is called the core distance of object p, written as core-distance (p), if object p is not the core object, then the core object of p is UN-DEFINED. In fact core-distance (p) is the minimum ε to make object p become a core object, or else core- distance (p) is UNDEFINED.

Definition 12) Reachable Distance (RD)

In the object se of the database O, reachable distance of object p to object q is the larger one of core-distance (q) and the distance from object p and object q, written as reachable-distance (p,q). If q is not the core object, then reachability-distance (p,q) is UNDEFINED. In fact, rea-chable-distance (p,q) is the minimum direct-density rea-chable distance from object p to object q. Thus, this rea-chable distance cannot be smaller than the core-distance (q), or else there's no direct density-reachable object to object q. The reachable distance of object p to core object is related to the core distance of the core object. The abo-ve two definitions are as shown as **Figure 3**.
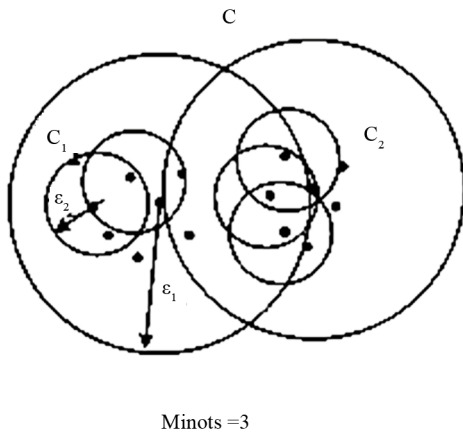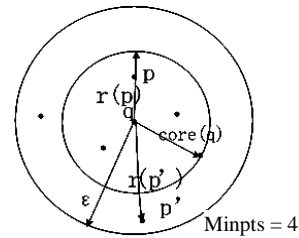

Figure 2. Global distance and local distance.

Minots =3

Figure 3. Core distance of object q and reachable distance of object q to object p.

Given the setting parameters Minpts and the initial distance ε, the procedure of OPTICS clustering algorithm can be described as follows:

Step1: Select an unclustered object p from the object set O of the database, and mark this object as clustered.

Step2: According to the parameter ε, get the object p's neighbor object set, and set core-distance (p) and reachable-distance (q) according to the neighbor set.

Step3: If core-distance (p) is UNDEFINED, namely this object is the core object, then add neighbors of object p into the priority queue.

Step4: If the priority queue is empty, then go on to the Step 5, or else select the minimum reachability-distance object q in the queue, get the neighbor set of object q, set the core-distance (q) and reachable-distance (q) according to the neighbor set. If object q is the core object, then add the neighbors of object q into the priority queue, re-peat Step 4.

Step 5: If all the objects in object set O have been marked as clustered, then go to the Step 6, or else to Step 1.

Step 6: Output the sorting result and clustering result according to the core distance and reachable distance of each object, then end this algorithm.

The major advantage of OPTICS algorithms is: it's not sensitive to the parameters Minpts and ε, that is, Minpts and ε do not influence the distribution of the reachable distance, the parameters can be selected randomly at a large scale without affecting the distribution of reachable distance and the inner clustering structure of object set. [8].Unlike DBSCAN algorithm, OPTICS algorithm can employ heuristic to estimate parameter ε and Minpts ac-cording to experience [9].

3.2. Design of Scientific Communities Found System

The System of scientific communities found based on citation network contains three independent parts: data preprocessing sub-system, communities mining sub-sys-tem, graphic demonstration sub-system. Shown as **Figure 4**, these three sub-systems can be divided into sev-eral independent function modules. Through transferring data, the three sub-systems are combined into a complex system.

4. Experiment and Analysis

In the experiment, SMARTY + PHP + MYSQL scheme is used to construct the model, in which SMARTY template is used for visualization display, PHP is used for processing data and controlling the procedure, and MYSQL is used for data storage.

Initial input data of the experiment come from the paper set of Wanfang Date (http://www.Wanfangdata.com. cn/). These initial data is acquired in HTML source code format, and processed by PHD program, to extract title, abstract keyword, author, references, publish date and other information of each paper. These information are stored in the MYSQL database.

Here, information of 22 papers and 31 authors is extracted from the original database, as shown as **Table 1** and **Table 2**. We can construct the citation network, in which each vertex represents each paper, named by the paper_id, and each edge represents the citation relationship between two papers.

Based on the citation network structure shown in **Figure 5**, we use Formular (1) to compute RPR between any two papers and get the RPR matrix of citation network. Similarly, we use Formulas (2) and (3) to compute the BCR matrix and CCR matrix [10]. Then according the Formula (4), we can get PR matrix. Furthermore, we can get AR matrix according Formula (5) and Formula (6), as shown as **Figure 6**. After computing the AR matrix. we can get the distance between any two authors and cluster the author objects.

In the experiment, we set $\varepsilon = 5$, Minpts = 4, the author object cluster order after data clustering is: 1, 2, 28, 29, 3, 24, 25, 4, 5, 6, 12, 7, 13, 26, 27, 9, 8, 20, 19, 11, 15, 14, 16, 10, 30, 31, 17, 18, 21, 22, 23, in which the numbers is id of the author object in the database). When we Set $\varepsilon = 3.0$, 2, 28, 29 belong to one cluster; 24, 25 belong to one cluster; 6, 12, 7, 13, 26, 27, 9, 8, 20, 19, 11, 15, 14, 16, 10, 30, 31, 17, 18 belong to another cluster; the others are noise. That means that author 2, 28, and 29 form a scientific community; author 24 and 25 form another community; 6, 12, 7, 13, 26, 27, 9, 8, 20, 19, 11, 15,
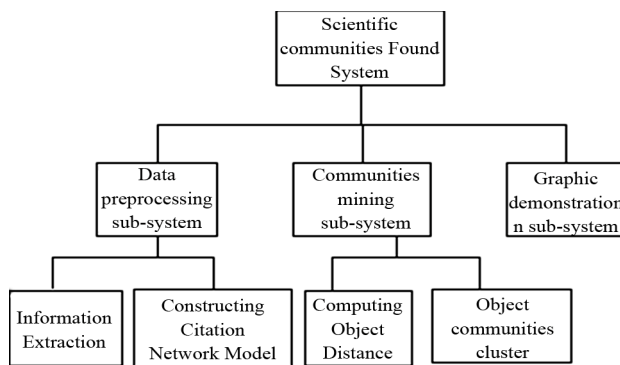

Figure 4. Scientific communities found system.

Table 1. Papers' information extracted from the original database.

Paper_id	title	author_name	refer id
1	Real-Time Fur Representation with Non-Uniform Layers of Textured Slices	Gang Yang; Hanqiu Sun;	2;
2	GPU-Based Realistic Fur Rendering	Gang Yang; Hanqiu Sun;	
3	3-Dimensional Visual simulation of Koch curve ased on the soil pore	Goming Wang; Xinyu Guo;	
4	Simulation system of growth based on corn process	Xinyu Guo; Chunjiang Zhao	
5	Spatial Hashing Based collision detection algorithm for virtual plants	Changfeng Li; Xinyu Guo;	19;
6	Soil Visual simulation study based on particle system	Goming Wang; Xinyu Guo;	
7	Realtime Simulation of Rain and Snow based on particle system and OpenGL	Liming Xu; Yuming Jiang;	
8	Virtual Walkthrough Collision Detection Based on Object-oriented Octree	Goming Wang; Xinyu Guo;	
9	Virtual maize model plant morphological constructing based on organ biomass accumulation	Youhong Song; Baoguo Li;	
10	Three-dimensional visualization of maize based on growth models	Xinyu Guo; Chunjiang Zhao	22;4
11	Virtual maize model biomass partitioning based o plant topological structure	Youhong Song; Baoguo Li;	
12	Overview of camera calibration for computer vision	Maolin Qiu; Songde Ma;	
13	Cartoon fur textures synthesis and replacement in images and videos	Shaohui Jiao; Gang Yang	
14	Modeling and realization of roaming virtual scene	Liming Xu; Yuming Jiang;	2;
15	Study on optimizing technology of particle system	Goming Wang; Xinyu Guo;	
16	Real-time simulation of snow falling and accumulation in agricultural scene	Wei Wei; Xinyu Guo;	7;
17	View-dependent greenhouse scene modeling and interactive walk through	Wei Wei; Xinyu Guo;	14;
18	Research on information standard for digital agriculture	Xinyu Guo; Chunjiang Zhao	
19	Research of Review virtual plant	Miao Guo; Baoguo Li;	
20	An adaptively deformable model for Web-based Virtual Surgery	Jishi Cai; Hanqiu Sun;	
21	Design and implementation of three-dimensional geometric morphological modeling and visualization system for maize	Xinyu Guo; Chunjiang Zhao	19;10;4
22	Study on the 3-Dimensional Visualization of Leaf morphological formation in corn	Xinyu Guo; Xuyang Deng	19;

IJIS

Table 2. Part authors' information extracted from the database.

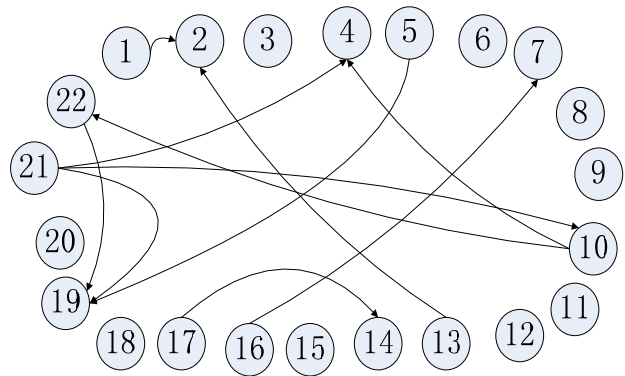| author_id	author_name	paper_id list
1	Gang Yang	1; 2; 13;
2	Qiuhan Sun	1; 2; 20;
3	Wencheng Wang	1; 2;
4	Enhua Wu	1; 2; 13;
5	Gongming Wang	3; 6; 8; 15;
6	Xinyu Guo	3; 4; 6; 8; 10; 15; 16; 17; 18; 21; 22;
7	Chunjiang Zhao	3; 4; 6; 8; 10; 15; 16; 17; 18; 21; 22;
8	Jihua Wang	4; 6; 8; 15; 21; 22;
9	Changfeng Li	5;
10	Shenglian Lu	5; 6; 17;
11	Yueying Yang	5;
12	Liming Xu	7; 14;
13	Yuming Jiang	7; 14;
14	Yonghong Song	9; 11;
15	Miao Guo	9; 11; 19;

Figure 5. Citation network structure.

14, 16, 10, 30, 31, 17, 18 can form a large scientific community; the other authors are scatter points.

5. Conclusions

This article researched on the structure of the citation network based on the citation path [11], defined the association relations between papers to transfer the scientific communities found model in the database to a clustering issue of points in the space. Besides, by using OPTICS algorithm, there's no need for users to concern about the selection of the parameters and clusters of any shape can be found effectively in a more practical way [12].

作者id	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	0	4	3.58	4.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.08	1.08	0	0	0	0	0	0
2	3.6	0	3	3.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.58	0.58	0	0	0.7	0.7	0	0
3	3.6	3	0	3.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.58	0.58	0	0	0	0	0	0
4	4.7	4	3.58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.08	1.08	0	0	0	0	0	0
5	0	0	0	0	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	2	0	20.1	9.27	1.4	2.4	1.43	0.67	0.7	0	1.87	1.87	3.889	3.89	11.2	7.444	0	0	0	0	0	1	0.67	0	0	3.56	3.71
7	0	0	0	0	2	20.1	0	9.27	1.4	2.4	1.43	0.67	0.7	0	1.87	1.87	3.889	3.89	11.2	7.444	0	0	0	0	0	1	0.67	0	0	3.56	3.71
8	0	0	0	0	2	9.27	9.27	0	0.7	0.7	0.7	0	0	0	0.78	0.78	2.556	2.56	4.4	3.444	0	0	0	0	0	0	0	0	0	0.89	0.96
9	0	0	0	0	0	1.43	1.43	0.7	0	0.4	0.4	0	0	0	0.13	0.13	0.333	0.33	1.03	0.833	0	0	0	0	0	0	0	0	0	0.5	0.2
10	0	0	0	0	0	2.43	2.43	0.7	0.4	0	0.4	0.67	0.7	0	0.13	0.13	0.333	0.33	1.03	0.833	0	0	0	0	0	1	0	0	0	0.5	0.2
11	0	0	0	0	0	1.43	1.43	0.7	0.4	0.4	0	0	0	0	0.13	0.13	0.333	0.33	1.03	0.833	0	0	0	0	0	0	0	0	0	0.5	0.2
12	0	0	0	0	0	0.67	0.67	0	0	0.7	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0	0	0	0
13	0	0	0	0	0	0.67	0.67	0	0	0.7	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.33	1.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	1.87	1.87	0.78	0.1	0.1	0.13	0	0	1.3	0	2.33	0.25	0.25	1.03	0.472	0	0	0	0	0	0	0	0	0	0.22	0.56
16	0	0	0	0	0	1.87	1.87	0.78	0.1	0.1	0.13	0	0	1.3	2.33	0	0.25	0.25	1.03	0.472	0	0	0	0	0	0	0	0	0	0.22	0.56
17	0	0	0	0	0	3.89	3.89	2.56	0.3	0.3	0.33	0	0	0	0.25	0.25	0	0.33	2.89	2	0	0	0	0	0	0	0	0	0	1.67	0.89
18	0	0	0	0	0	3.89	3.89	2.56	0.3	0.3	0.33	0	0	0	0.25	0.25	0.333	0	2.89	2	0	0	0	0	0	0	0	0	0	1.67	0.89
19	0	0	0	0	0	11.2	11.2	4.4	1	1	1.03	0	0	0	1.03	1.03	2.889	2.89	0	5.444	0	0	0	0	0	0	0	0	0	2.56	1.84
20	0	0	0	0	0	7.44	7.44	3.44	0.8	0.8	0.83	0	0	0	0.47	0.47	2	2	5.44	0	0	0	0	0	0	0	0	0	0	2	1.44
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0.7	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0.7	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0.7	0	0	0	0	0	0	0	0	0
24	1.1	1	0.58	1.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0
25	1.1	1	0.58	1.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0
26	0	0	0	0	0	1	1	0	0	1	0	0.67	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0.67	0.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0
29	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0	0
30	0	0	0	0	0	3.56	3.56	0.89	0.5	0.5	0.5	0	0	0	0.22	0.22	1.667	1.67	2.56	2	0	0	0	0	0	0	0	0	0	0	0.56
31	0	0	0	0	0	3.71	3.71	0.96	0.2	0.2	0.2	0	0	0	0.56	0.56	0.889	0.89	1.84	1.444	0	0	0	0	0	0	0	0	0	0.56	0

Figure 6. AR matrix.

In order to test and analyze conveniently, the number of selected papers is not large and all of them are in Chinese. This article hypothesis there's no fault reference or repeated reference, while in fact, these problems are quite likely to happen. So our future work is to improve the model, doing research on how to deal with those issues and filter the citation path accurately.

6. Acknowledgements

REFERENCES

[1] X. M. Liu, J. Bollen and M. Nelson, "Co-Authorship Networks in the Digital Library Research Community," *Information Processing and Management*, Vol. 41, No. 5, 2005, pp. 1462-1480. doi:10.1016/j.ipm.2005.03.012

[2] R. Ichise, H. Takeda and T. Muraki, "Research Community Mining with Topic Identificatio*n,"* Proceedings of the Information Visualization, London, 2006, pp. 276-281.

[3] H. X. Lin, "Scientific Group Mining Based on the Citation Network," Master Dissertation, Fudan University, Shanghai, 2009.

[4] N. Ma and J. C. Guan, "Survey on the Mining Algorithm Based on the Network Structure," *Informatics Disciple*, Vol, 13, No. 1, 2008, pp. 3-14.

[5] K. Börner, L. Dall'Asta, W. Ke and A. VeSPignani, "Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship teams," *Complexity*, Vol. 10, No. 4, 2005, pp. 57-67. doi:10.1002/cplx.20078

[6] M. J. Bommarito, D. M. Katz, J. L. Zelner and J. H. Fowler," Distance Measures for Dynamic Citation Networks," Elsevier, Amsterdam, Vol. 289, No. 19, 2010, pp. 4201-4208.

[7] Y. Takeda and Y. Kajikawa, "Tracking Modularity in Citation Networks," *Scientometrics*, Vol. 82, No. 3, 2010, pp. 783-792. doi:10.1007/s11192-010-0158-z

[8] G. Z. Wang and G. L. Wang, "Improved Quick DBSCAN Algorithm," *Journal of Computer Applications*, Vol. 26, No. 15, 2009, pp. 2364-2373.

[9] S. K. Zhou, A. Y. Zhou and J. Cao, "DBSCAN Algorithm Based on Data Partition," *Journal of Computer Research and Development*, Vol. 2, No. 2, 2000, pp. 169-194.

[10] H.-P. Kriege and M. Pfeifle, "Density-Based Clustering of Uncertain Data," *Proceedings of the* 11*th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, New York, 2005. doi:10.1145/1081870.1081955

[11] J. Liu and X. S. Ma, "Improved DBSCAN Clustering Algorithm's Study and Application," *Communication and Computer*, China, 2008.

[12] Y. L. Zeng, H. B. Xu and S. Bai, "Improved OPTICS Algorithm and Its Application in Text Clustering," *Journal of Chinese Information Processing*, Vol. 9, No. 4, 2008, pp. 1483-1492.