Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 4, Issue 3: 68-76, DOI: 10.25236/AJCIS.2021.040311

Doctor recommendation via community detection

Li Guo*, Chenning Du

Dalian Polytechnic University, Dalian, China
**Corresponding author: Guoli@dlpu.edu.cn*

Abstract: *With the development of socioeconomic and computer science and technology, people have paid more and more attention to physical health and living convenience, network precision medicine enters people's lives. Because that teamwork is needed in the medical process, how to recommend a doctor to form a team becomes important. Instead of using recommendation algorithm, the paper recommends doctors based on community detection. Firstly, 37144 pairs of doctors with labels based on thesis cooperation relationship were obtained through web crawler, secondly, find the connected subgraph with the union check set, thirdly, create the graph with the doctors as nodes, the cooperation relationship between doctors as the edges, and cut the graph with the algorithm of spectral clustering until the connected graphs have approximately five layers of relations. Then take the number of cooperative papers as the weight of edges. Recommend a doctor with the doctors who are in at least the third layer and with a higher weight score in the same community. At the same time, the paper suggests the doctors of the same community make the team. From the perspective of community detection, this paper solves the problems of doctor recommendation and team building, which results in more ideal results. And the community visualization paves the way for the recommendation system to be completed in the future.*

Keywords: *Community detection, Doctor recommendation, Spectral clustering*

1. Introduction

At present, the Internet has changed the traditional service mode of seeking medical advice and consulting, the spread of medical information is more convenient, and the connection between doctors and patients is more closed. As a behavioral agent in the field of medical health, doctors can even get rid of the bondage of medical institutions at all levels and play an increasingly important role in Internet health care. The analysis of doctor relationship network is of great significance for research on the mode of Internet medical treatment in the new form. The doctor relationship network has the characteristics of complexity and diversity, this paper excavates the internal relationship between doctors from the perspective of paper cooperation, and based on this, provides suggestions for the composition of doctor teams to improve the quality of internal medical service. Moreover, research on community detection is not only the hot spot and the important study direction in the field of a complex network but also the hottest topic of application of data mining in complex network [1]. Especially in the social network, a real social group can be displayed by interest, occupation, region, and background. In this way, we can carry out character analysis, career recommendation, circle recommendation, friend recommendation, etc.

To detect the structure of the doctors' cooperative relationship, it is necessary to find the related reference and determine the edge weight of the network. For example, [2] uses a number of times where a retweet is observed between two Twitter users as edge weight of a Twitter network. Thus, the related reference is retweet. The phone calls between customers [3] or the exchanged e-mails within an institution [4] and so on are also related reference by which weighted graphs can be created. Considering the academic communication of the doctors can reflect their specialty of the field better, and the communities based on this is also more representative and stable, the paper cooperation is chosen to be the related reference, and the number of the cooperative papers to be the edge weight.

After the original network is displayed, some Unicom-subgraphs with hundreds of nodes, even one subgraph with more than ten thousand edges is shown. According to the five-degree segmentation theory, only five layers of characters following the core characters can be seen the effective social networking, therefore, extracting the Unicom-subgraphs and conducting community detection on the subgraphs don’t satisfy the five-degree segmentation theory is very important.

As for the problem of extracting the Unicom-subsets, the paper applies the algorithm of Union-Find, which is specially used to deal with problems such as a combination of disjoint sets and quick query in a set. As for the problem of selecting the community detection algorithm, there are two points, one is which algorithm has the best effect on our dataset, the other one is how to evaluate the algorithm's effects. We tested many types of algorithms on our dataset to conduct community detection, including FN (Fast Newman) which is based on the idea of aggregation, GN (Grivan-Newman) which is based on the idea of split, walk trap which is based on information theory and SC (Spectral Clustering) which is based on spectrum analysis. We chose modularity Q, which is an important index function to measure the division of the network, it can quantify the advantages and disadvantages of the community detection results [5], even as the evaluating indicator to evaluate the results of the algorithms. SC finally shows the best effect.

Furthermore, this paper’s community detection is unsupervised, terminate conditions become an important subject, many algorithms such as GN and FN use maximum modularity as the terminate condition, in this paper, except for the maximum modularity, we also add the five-degree segmentation theory to be the terminate condition.

The remaining part of this paper is organized as follows. In Section 2, we introduce the algorithm of SC proposed in this paper. The evaluating indicator, experiment results of the four algorithms on the test dataset and the final detected network are presented in Section 3. Section 4 summarizes the related works to the community detection and its evaluating indicator. Finally, in Section 5, we conclude the paper and suggest further study of this work.

2. Methodology

The framework of this paper`s methodology is shown as Fig.1. Which includes the steps of data collection, data processing, and community detection and construct plots. For the step of community detection, if the classified relationship network doesn't satisfy the five-degree segmentation theory, there will be an iteration until the number of layers is no more than five.

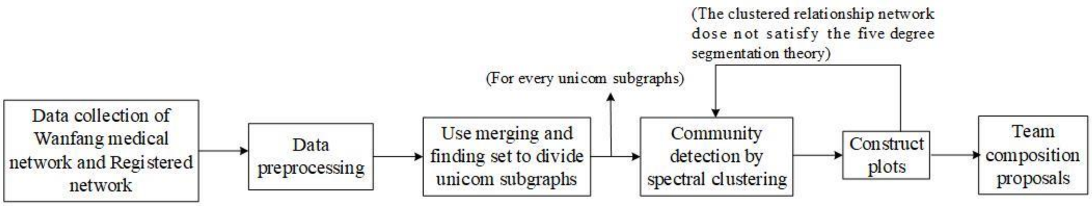

Figure 1 Framework

2.1 Basic Concepts

The Union-Find is a tree-type data structure. It is used to deal with the merge and query problems of some disjoint sets. The refinement of Union-Find is that the tree is used to represent the set, and the root node of each tree is the representative element [10] of the corresponding set of the tree. There are three main operations in the algorithm of Union-Find: Initializing the set, finding the root node, and merging the collection. The initialization operation is to initialize the Union-Find set to a forest, and each tree in the forest contains only one root node. The operation of finding the root node is to find the representation element for each set while the operation of merging is to merge the two sets into one set. In this paper, we use Union-Find to merge the ID sets in which every set contains two id numbers presenting the two doctors who have cooperation with each other into one set, that is, one connected subgraph. Then the connected subgraphs are divided by spectra clustering, after that, we use Gephi, an open source, free cross-platform complex network analysis software based on JVM, which is mainly used for various networks and complex systems, interactive visualization and exploration of open source tools for dynamic and layered graphs to construct the network plots.

With the development of the network, the relationship between people is becoming more and more close [11]. The social networking giant Facebook and the University of Milan jointly announced their latest research on the six-degree separation theory: The average number of people between two independent persons in the world is 4.74. Compared to the result of 6 people of Stanley Milgram's experiment conducted in 1967, 1.26 people have been reduced. Therefore, when building the character relationship network, we only need to collect five layers of characters following the core characters, which reduces the depth of the relationship network, and makes the classified relationship network more practical. So, if the classified relationship networks have more than five layers under the core character, we will iteratively conduct the community detection, and if the network doesn't have a core character, we will also iteratively conduct the community detection until the subnetwork satisfies the five-degree segmentation theory.

2.2 Data acquisition and preprocessing

The distributed network crawler system is improved from the traditionally centralized network crawler, whose working principle is similar to the centralized web crawler. The distributed web crawler system is regarded as several centralized web crawlers, which are connected by a certain way of communication and organization to coordinate the crawler system. Two kinds of data sets are preliminarily grabbed through the web crawler. Dataset A is the information about doctors' cooperative papers with a total of 3017526 pieces of data captured from the Wanfang Medical Network. It mainly includes doctor ID, co-author ID, and the number of cooperative papers. Dataset B has a total of 29474 pieces of data, which is the doctors' team information of the National Tertiary Hospital captured from the Guahao Network, including doctors' names, hospitals' names, team names, etc.

According to the doctor ID in the dataset A, the doctors' names and hospital are obtained from the information table of Wanfang medical database. Save the information to the Doctor file, and get the dataset C, which mainly includes the doctor ID, the collaborator ID, the number of cooperative papers, the names of the doctors, and the names of the hospitals.

Associate dataset B with dataset C and there are two association rules: One is that the names of the doctors of the two datasets are the same, the other one is that the names of the hospitals of the two datasets are the same. By the above rules, the information of the doctors' cooperative papers and the information of the doctors' teams can be integrated together. Save the information in the Final file to get the final data.

2.3 The algorithm of spectral clustering

Spectral clustering is a clustering algorithm based on graph theory [9], clustering the eigenvectors of the Laplace matrix of the sample data. It has the advantages of processing data sets of arbitrary shape and easy to execute [6]. Recently, it has been used more widely, such as image segmentation [7], community mining in complex networks [8], etc, even though until now rarely used in medical-related community detection. However, the paper found it applied to the data set of medical better than other algorithms to do community detection.

A graph consists of points and lines between points. A point represents a thing, and a line represents a kind of relationship between the two things, which is also called weight. The map is divided into several subgraphs, each subgraph has no intersection, and the sum of the weight of the line which is cut off between the subgraphs is called the loss function. Spectral clustering realizes graph partitioning by minimizing the loss function. Let $G(V,E)$ represents the graph, $V(v_1,v_2,\dots,v_n)$ represents the set of the points, E is said to be the edge set. w_{ij} represents the weight between v_i and v_j. Suppose $G(V,E)$ is divided into two subgraphs of G_1 and G_2, define $q=[q_1,q_2,\dots,q_n]$ as a n-dimensional vector, which is used to represent the partition scheme.

$$\mathrm{q_i} = \begin{cases} c_1 & i \in G_1 \\ c_2 & i \in G_2' \end{cases} \tag{1}$$

The loss function can be defined as:

$$\mathrm{Cut}(\mathrm{G_1},\mathrm{G_2}) = \sum_{i\in G_1, j\in G_2} w_{ij} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(q_i-q_j)^2}{2(c_1-c_2)^2}, \tag{2}$$

in which: W is weight matrix and D is diagonal matrix.

$$\mathrm{D_{ii}} = \sum_{j=1}^{n} w_{ij}, \tag{3}$$

The definition of the Laplace matrix $L=D\text{-}W$, minimizes the loss function problem into the minimization of q^TLq.

2.4 The application of spectral clustering in community detection

The community discovery algorithm based on spectral clustering considers the elements in the community as the point set *V* of the graph, the interconnections between elements can be seen as the line *E* that connects the vertices of the graph, based on the correlation's degree of tightness between elements, the weighted undirected graph *G(V, E)* is constructed. Thus the problem of clustering can be transformed into the problem of graph division. The optimal partition criterion based on graph theory is that the inner similarity of subgraphs is maximum and the similarity between subgraphs is minimum. In this way, community detection is carried out, and the concrete steps are as follows :

Step1 Mapping the community to be divided into weighted undirected graph *G(V, E)*.

Step2 Calculate the similarity matrix *W* and Laplace matrix *L*.

Step3 Making characteristic decomposition of the Laplace matrix *L* to obtain the eigenvalues and eigenvectors.

Step4 K-means or other classical clustering algorithms are used to cluster eigenvectors in the eigenvector space, completing the community detection.

3. Experiments

There are two original datasets, Dataset A is the information about doctors' cooperative papers with a total of 3017526 pieces of data captured from the Wanfang Medical Network, which mainly includes doctor ID, co-author ID, and the number of cooperative papers. Dataset B has a total of 29474 pieces of data, which is the doctors' team information of the National Tertiary Hospital captured from the Guahao Network, including doctors' names, hospitals' names, team names, etc, as shown in 3.1. After data processing, we got the data set of 37144 pairs of cooperative doctors, including doctor ID, co-author ID, the number of cooperative papers, doctors' names, hospitals' names, and team names. We used doctors as the nodes in the graph, the cooperative relationship of doctors as the edges, and the number of cooperative papers as the weight of the corresponding edge to construct the graph. Then these data are processed by Union-Find, and resulting in 885 connected subgraphs. We chose the subgraph with the most nodes of 187 as the baseline's data set and used GN(Grivan-Newman)[12-13], FN(Fast Newman) [14] and Walktrap [15] as the contrast algorithms to SC(Spectral Clustering).

3.1 The Evaluating Indicator of Modularity

Modularity, which is also known as the value of modularity, is a commonly used method to measure the strength of the community's network structure, and it [16] was first proposed by Mark NewMan. The definition of modularity is:

$$Q = \frac{1}{2m} \times \sum_{ij} \left[A_{ij} - \frac{k_i \times k_j}{2m} \right] \Delta(C_i, C_j), \tag{4}$$

The size of modularity mainly depends on C, which is the community distribution of the nodes in the network. Namely how the network is divided through community detection, can be used to quantitatively measure the quality of community detection for the network. And the more its value is closed to 1, the stronger the intensity of community structure is, and the better the detection quality is. Therefore, the optimal network community partition can be obtained by maximizing the modularity degree Q.

3.2 Experimental Results

GN is an algorithm of community detection with disintegrated type. According to the characteristics of high cohesion in the community and low cohesion among communities, the algorithm gradually removes the edges between communities and achieves relatively cohesive community structure. The algorithm uses the concept of the edge betweenness to detect the location of the edges, and the edge betweenness of an edge is defined as the number of the shortest paths between all nodes of the network, and the paths must contain the edge. By definition, if one edge connects two communities, then the number of the shortest paths passing the edge will be the maximum, and the paths are between the two communities' nodes, the corresponding edge betweenness will be the largest. If the edge is deleted, then the separated two communities will occur. Based on this idea, GN calculates the shortest paths of the current network repeatedly, calculates the edge betweenness of every edge, and deletes the edge with the largest edge betweenness. At last, under certain conditions, the community structure of the network can be obtained by stopping the algorithm.

The algorithm of FN is a kind of fast mining algorithm based on the modularity, belonging to the

ISSN 2616-5775 Vol. 4, Issue 3: 68-76, DOI: 10.25236/AJCIS.2021.040311

agglomerative algorithm, which is a kind of greedy algorithm. According to the modularity Q, two connected communities are traversed to find the combination of maximum or minimum increment, merged into a new community, and a hierarchy tree is constructed. The algorithm process is defined as:

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2\,(e_{\{ij\}} - a_i a_j) \tag{5}$$

Every time after the combination, for the new communities, its symmetric matrix *e* is renewed. The number of maximum connection steps of the algorithm is *n-1*, the hierarchy tree can be constructed then use the modularity degree Q to select the optimal truncation surface to conduct the community detection.

The basic idea of a walk trap module is to traverse a graph from one or a series of nodes. At any vertex, the more ergodic will travel to the vertex next to this vertex by the probability of $1-\alpha$, with the probability of α randomly jumping to any vertex of the graph, then α is called the probability of jump occurrence. A probability distribution is obtained after every walk, the probability distribution characterizes the probability that every vertex is accessed. This probability distribution is used as the input of the next walk trap and repeatedly iterate this process. When a certain of precondition is satisfied, the probability distribution tends to be converging. After convergence, a stable probability distribution can be obtained, that is the community structure is obtained.

For the test dataset, through the experiment, we found that the maximum modularity of every algorithm occurs when the community is divided into 30 categories. So with in 30 categories, every time the community is divided, we calculated the modularities of the SC, GN, FN, and Walktrap, the results are shown in Fig 2.

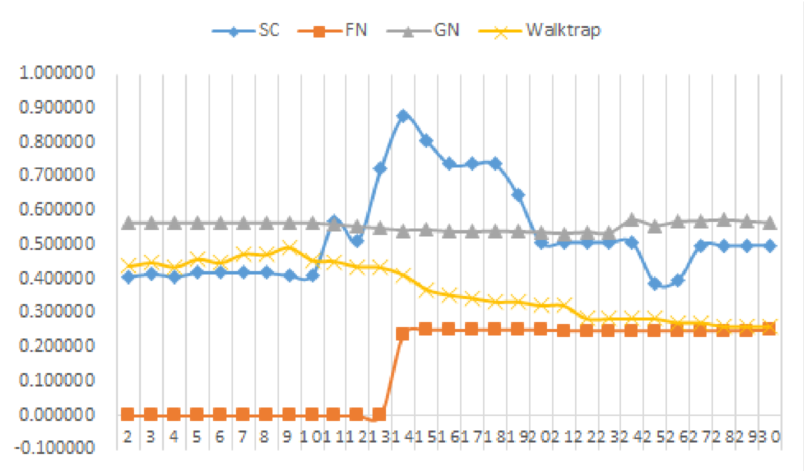

Figure 2 Experimental results

Now we can see that for the dataset, when the number of divided communities changing from 2 to 11, GN and Walktrap has little larger modularity than SC, and from 20 to 30, GN also shows a little better effect than SC. But from 12 to 19, SC always shows the best effect and gets the largest modularity of 0.88. Generally, we will choose the status of the community structure when its modularity gets the largest. Moreover, this test set is quite representative, so we chose SC to do the community detection in the paper.

3.3 Network Visualization

For every graph of the 885 connected subgraphs obtained through Union-Find, we iteratively use spectral clustering to conduct clustering and divide communities until the network the subgraph satisfies the five-degree segmentation theory, for every time's clustering, we chose the result that the community structure has the largest modularity, and finally use Gephi to visualize the network shown in Fig 3.

ISSN 2616-5775 Vol. 4, Issue 3: 68-76, DOI: 10.25236/AJCIS.2021.040311

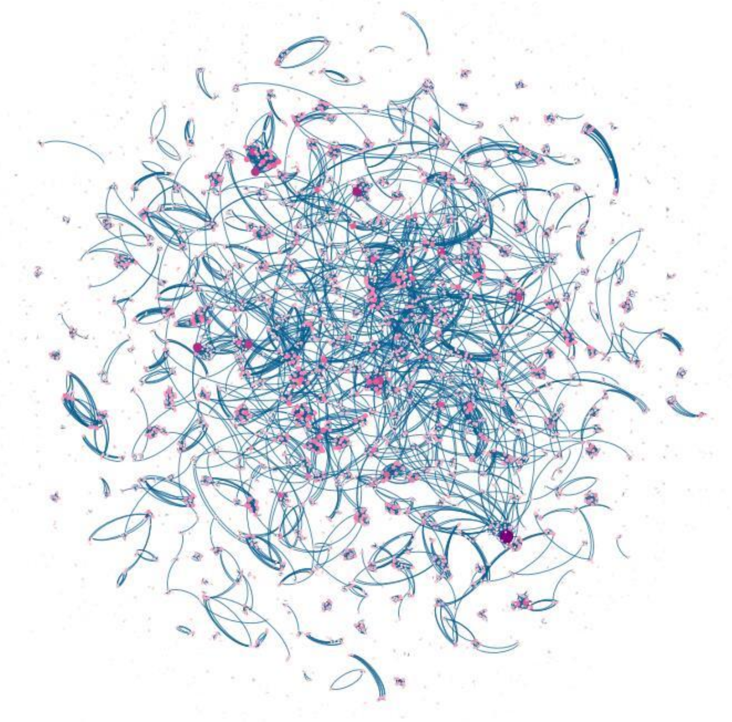

Figure 3 Network visualization

Compared with the original network, which has 37144 edge connections, our final network with 9255 edges cut 27889 edges to make the network more efficient and the structure stable. According to the final network, we can recommend the doctors in the same cluster to make the team. Take one cluster shown in Fig 4 as an example. The nodes present for doctors, edges for the connections between doctors, that's the number of their coauthor papers and the more they cooperated, the greater is the weight, the thicker is the connection line. According to the network, Rihua Jiang and Jianxin Xia both have direct cooperation with the other six persons, however, Rihua Jiang, Zhongmin Jiang and Mingji Zhu from the same hospital obviously have close connections with each other, moreover, the network obeys five-degree segmentation theory, so people in the network are recommended to build up a team and Rihua Jiang to be the team leader.

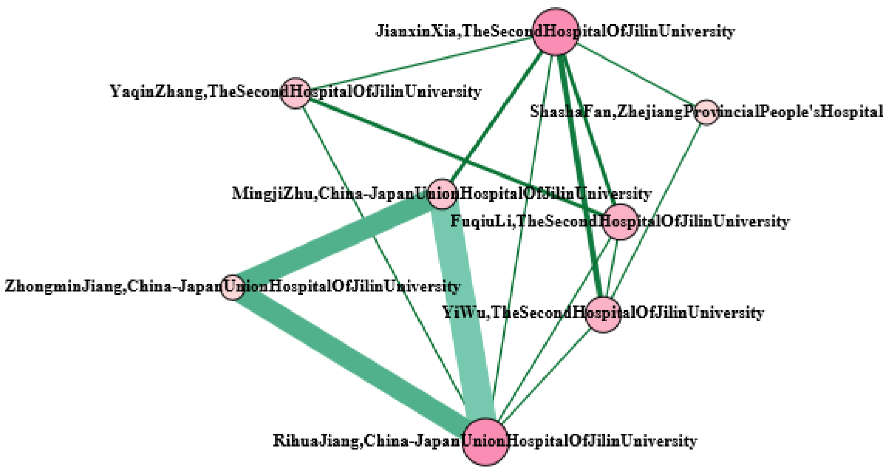

Figure 4 Case study

ISSN 2616-5775 Vol. 4, Issue 3: 68-76, DOI: 10.25236/AJCIS.2021.040311

4. Related Works

The related work can be grouped into two categories. The first category is the studying of community detection, the second category related the evaluating indicator of non-overlapping community detection algorithms.

4.1 Community Detection

Studying the community in the network plays a crucial role in understanding the structure and function of the whole network, and helps us to analyze and predict the interactions among all the elements in the network. Moreover based on spectral graph theory, spectral clustering has attracted much attention in recent years because of its many advantages.

In order to make it possible to analyze large network systems, [17-19] et al use and compare existed algorithms, improve the traditional community detection algorithms or put forward new methods, finally testing them on several specific networks. [20-21] et al discusses the topic of community detection in the context of social media. In [22], the authors present a multi-stage algorithm based on local-clustering, and they apply this to the YouTube video graph to generate named clusters of videos with coherent content, which is a first of its kind. [23] attempts a thorough exposition of the graphs representing real systems, from the definition of the main elements of the problem, to the presentation of most developed methods, compares them from the key issues. [24-25] conduct the graph partition by adapting the Laplacian spectral partitioning method, the algorithm is tested on computer-generated and real-world networks, showing good performance. In [26], the author proposed a hybrid collaborative filtering method to solve the cold start and data sparsity problems in CF. Because now the detection of hidden community structure has become an important subject, [27] connects modularity-based methods with correlation analysis, proposing a framework in the correlation analysis research area to advance community detection.

Different from the research mentioned above, which focus on improving the traditional algorithms especially in social network area or processing graphs, we use the algorithm of spectral clustering, which shows the best performance on the dataset to conduct the community detection on the doctor's cooperative network in the area of big data for medical treatment.

4.2 The Evaluating Indicator

Community detection algorithms have become a research hot spot in the field of complex network. With the development of research, a large number of algorithms have emerged. How to evaluate the performance of the algorithm becomes a topic.

For the non-overlapping communities, [28] compares and analyses the evaluating indicators of community detection algorithms and finds that modularity Q has the maximum scope of application, NMI applies to a small number of communities, and ARI applies to a large number of communities. There are also some researches focus on actual technologies to select the most suited algorithm according to the properties of the network under consideration, for example [29], which also use the hybrid parameter as an easily measurable indicator of finding the range of reliability of different algorithms. [30] proposes methods and metrics for evaluating graph clustering results to be used in not only social networks, but also biological networks, computer vision, and image processing, etc. Moreover, some new algorithms of community detection are proposed based on modularity, [31] puts forward a method based on modularity and an improved genetic algorithm (MIGA), [32] proposes Fine-tuned based on modularity (Q) or modularity density (Qds). [33] employs a novel cross-dimension network validation (CDNV) procedure to compare the performance of different methods. Moreover, some papers optimize existed evaluating indicator, then use it to the algorithms, or put forward a novel evaluation model, [34] and [35] are in this case. [36] recommends applying two types of evaluation, one with the high performance, the other connecting topological properties of the communities, which are considered as complementary to perform a complete and accurate assessment.

From above, we can see there occurs many indicators to evaluate specific or non-specific networks, which are almost based on modularity. However, despite their initial success, they don't connect the evaluating indicator with spectral clustering to conduct community detection. Thus, our work tries to evaluate and compare the community detection algorithms including spectral clustering.

ISSN 2616-5775 Vol. 4, Issue 3: 68-76, DOI: 10.25236/AJCIS.2021.040311

5. Conclusion

This paper conducts community detection on the social network of doctors. We built network with doctors as nodes, and the number of cooperative papers as the weight of the edges between doctors. However, the original graph has some big messy networks, indicating the necessity to do the community detection. Therefore, we ran algorithms on the test dataset, which is a Unicom subgraph of the original network. Modularity Q is chosen to be the evaluating indicator. The results show that the algorithm of SC has the highest modularity, meaning it has the best performance to detect the structure of communities on our dataset. We iteratively ran spectral clustering, every time chose the detected structure with the highest modularity until every divided community obeys the five-degree segmentation theory. Finally, we recommend the doctors of the same community, which means they cooperated closely or there is great possibility of close cooperation between them to make a medical group engage in medical or research activities together. In the future, we will try to improve the spectral cluster to promote the modularity further.

Acknowledgements

Funding: Supported by National Key R&D Program of China (2018YFC0116800), National Natural Science Foundation of China (No. 61772110), CERNET Innovation Project (NGII20170711).

Data used in the preparation of this article were obtained from Haoyisheng, Wanfang and Guahao Online database.

References

[1] Coscia M, Giannotti F, Pedreschi D. A classification for community discovery methods in complex networks [J]. Statistical Analysis and Data Mining: The ASA Data Science Journal, 2011, 4(5): 512-546.
[2] Youcef Abdelsadek, Kamel Chelghoum, Francine Herrmann, Imed Kacem, Benoît Otjacques. Community extraction and visualization in social networks applied to Twitter [J]. Information Sciences, 2017.
[3] V. Blondel, J. Guillaume, R. Lambiotte, E. Mech. Fast unfolding of communities in large networks J. Stat. Mech (2008), p. P10008
[4] N. Schlitter, T. Falkowski, J. Lssig. Dengraph-ho: density-based hierarchical community detection for explorative visual network analysis Springer (Ed.), Research and Development in Intelligent Systems XXVIII Incorporating Applications and Innovations in Intelligent Systems XIX Proceedings of AI-2011, the Thirty-first SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer, London (2011), pp. 283-296
[5] Y. Hu, J. Wu, Z. Di. Enhance the efficiency of heuristic algorithms for maximizing the modularity Q [J]. EPL (Europhysics Letters), 2009, 85(1).
[6] Cai Xiao-yan, Dai Guan-zhong, Yang Li-bin. Survey on sprctral clustering algorithms [J]. Computer science, 2008, 35(7):14-48.
[7] Zhu Z, Wang L. Initialization approach for fuzzy C-means algorithm for color image segmentation [J]. Application Research of Computers, 2015, 32(4):1257-1260.
[8] Liu Da-you, Jin Di, He Dong-xiao, Huang Jing, Yang Jian-ning, Yang Bo. Community mining in complex networks [J]. Journal of computer research and development, 2013, 50(10):2140-2154.
[9] Pan H F, Dong L, Chen J N, et al. Simulation of Shape Clustering Based on Spectral Graph Theory [J]. Journal of System Simulation, 2011, 23(7): 1409-1413.
[10] Stephen Alstrup, Mikkel Thorup, Inge Li Gørtz, Theis Rauhe, Uri Zwick. Disjoint Set Union with Constant Time Deletions [J]. ACM Transactions on Algorithms (TALG), 2014, 1(6):1-28.
[11] M.Girvan, M.E.J.Newman. Communitys structure in social and biological networks [J]. PNAS, 2002, (99):7821-7826.
[12] Girvan M, Newman M. Community structure in social and biological networks [J].Proc Natl Acad Sci USA, 2002, 99:7821-7826.
[13] Newman M, Girvan M. Finding and evaluating community structure in networks [J]. Phycical Review E, 2004, 69: 026113.
[14] Newman MEJ. Fast algorithm for detecting community structure in networks[J]. Phycical Review E, 2004, 69(6): 066133.
[15] Grady L. Random walks for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(11): 1768-1783.

ISSN 2616-5775 Vol. 4, Issue 3: 68-76, DOI: 10.25236/AJCIS.2021.040311

[16] Newman M E J. Fast algorithm for detecting community structure in networks [J]. Physical review E. 2004, 69 (6): 066133.
[17] Lancichinetti A, Fortunato S. Community detection algorithms: a comparative analysis [J]. Physical review E, 2009, 80(5): 056117.
[18] De Meo P, Ferrara E, Fiumara G, et al. Generalized louvain method for community detection in large networks[C]//Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on. IEEE, 2011: 88-93.
[19] Liu X, Murata T. Community detection in large-scale bipartite networks [J]. Information and Media Technologies, 2010, 5(1): 184-192.
[20] Papadopoulos S, Kompatsiaris Y, Vakali A, et al. Community detection in social media[J]. Data Mining and Knowledge Discovery, 2012, 24(3): 515-554.
[21] Planti éM, Crampes M. Survey on social community detection [M]//Social media retrieval. Springer, London, 2013: 65-85.
[22] Gargi U, Lu W, Mirrokni V S, et al. Large-Scale Community Detection on YouTube for Topic Discovery and Exploration[C]//ICWSM. 2011.
[23] Fortunato S. Community detection in graphs[J]. Physics reports, 2010, 486(3-5): 75-174.
[24] Newman M E J. Community detection and graph partitioning[J]. EPL (Europhysics Letters), 2013, 103(2): 28003.
[25] Newman M E J. Spectral methods for community detection and graph partitioning[J]. Physical Review E, 2013, 88(4): 042822.
[26] Chang Su,Yue Yu,Xianzhong Xie,Yukun Wang. Data Sensitive Recommendation Based On Community Detection [J]. Foundations of Computing and Decision Sciences,2015,40(2).
[27] Lian Duan,Yanchi Liu,W. Nick Street,Haibing Lu. Utilizing advances in correlation analysis for community structure detection [J]. Expert Systems With Applications, 2017,84.
[28] Lina Zhao, Hui Li.Evaluation index analysis of non- overlapping community detection algorithms[C].2014 Eighteenth Annual Conference on new technology and application for network of the network application branch of ACU,Kunming,Yunnan,ACU,Beijing,China,p.378-381.
[29] Yang Z, Algesheimer R, Tessone C J. A comparative analysis of community detection algorithms on artificial networks [J]. Scientific Reports, 2016, 6: 30750.
[30] Malliaros F D, Vazirgiannis M. Clustering and community detection in directed networks: A survey [J]. Physics Reports, 2013, 533(4): 95-142.
[31] Shang R, Bai J, Jiao L, et al. Community detection based on modularity and an improved genetic algorithm [J]. Physica A: Statistical Mechanics and its Applications, 2013, 392(5): 1215-1231.
[32] Chen M, Kuzmin K, Szymanski B K. Community detection via maximization of modularity and its variants[J]. IEEE Transactions on Computational Social Systems, 2014, 1(1): 46-65.
[33] Malliaros F D, Vazirgiannis M. Clustering and community detection in directed networks: A survey [J]. Physics Reports, 2013, 533(4): 95-142.
[34] Hanhua Chen,Hai Jin.The detection and evaluation of the community structure in the semantic peer-to-peer overlay network[J].Chinese Science: Information Science,2012,42(05):537-549.
[35] Wenwei Zhan,Jingke Xi,Zhixiao Wang.Hierarchical aggregation community detection algorithm based on similarity modularity[J].Journal of system simulation,2017,29(05):1028-1032+1040.
[36] Orman G K, Labatut V, Cherifi H. Comparative evaluation of community detection algorithms: a topological approach [J]. Journal of Statistical Mechanics: Theory and Experiment, 2012, 2012(08): P08001.