Detecting the Overlapping Community by Using Extract Method

Yuejiao Ma*, Yongjun Liu and Ji Tao
Huaneng Lancang River Hydropower Co., Ltd, Kunming City, China
*Corresponding Author Email: 13987625560@139.com

Abstract. The power overlapping communities are getting more and more attention in the complex network. In many networks, communities are not disjoint. It is very difficult to get complete information of the network, such as the WWW, the local community detection has been studied by researcher. In this paper, we applied the local community detection to detecting the overlapping communities. Based on the algorithm in [1]. In this paper, we make an improvement to let the source node selected in each time is fixed, and then the algorithm has a certain stability. Even though the experimental results have no great improvement, but our improvement makes the algorithm has a certain stability.

1. Introduction

Complex networks have been used to represent various kinds of system in the real world [1,2,3,4]. The important feature of complex networks is community structure [5, 6, 7, 8], i.e., the existence of nodes within a group are much more connected to each other nodes. There are more than the rest of the network. Communities detection often reflect important relationships among individuals (vertices). for example, there are physical connections between computers, friendships among people, and citations among research papers.

There are many community detection algorithms, which divide a network into disjoint clusters of vertices. However, many communities are not disjoint in plenty of network: for example, many people may be different social communities, which depend on their friends, parents, professions, friends, hobbies, etc[11]. Some algorithms [9-14] are able to detect these overlapping communities.

But, there are many real-world social network, WWW. These networks share the common feature. these complete structure of the network cannot be unavailable, because the entire network is too big and too dynamic[15]. And in some examples, complete information of the graph cannot be got, we only just can get local information, or some local information can be known. Therefore, the global community detection is not available. To solve the above problems, local community detection has been proposed. which only information of partial vertices is known, such as these algorithms [16-18]. The local community detection is to extract a local community with a certain source vertex.

In this paper, the local community detection is applied to detecting the overlapping communities. We made improvements based on [14]. [14] put forward this idea, but the community detected is unstable. Since choosing each source node is random, there is great uncertainty. In this paper, we made an improvement that choosing each source node is the same every time and then the detecting result at each time is the same.

This paper layout is as follows. Part II presented the definition of local community detection, the overlapping communities and reviewed in the preceding papers and algorithms. Part III describes our method. Part IV presents the experimental results and Part V gives the conclusion.
2. Problem Defined and Related Work.
In the graph, some nodes belong to two or more communities. See the Fig.1.

![Figure 1](image1.png)

Figure 1. The red nodes belong to a community, and the blue nodes belong to another community, and the two green nodes is belong to the two communities. Therefore, these two communities are overlapping communities and the two green nodes are overlapping nodes.

The definition of local community detection is starting from a source vertex i in the graph and extract a community contain i. See the Fig.2.

![Figure 2](image2.png)

Figure 2. The blue nodes are the extracted community, and the green nodes are neighbors of the community, and the red nodes.

3. Our Method
In 2.3, we reviewed the method in [14]. And we found a big problem that in the process of the algorithm, the source node choosing in each local community division is random which lead to the result of running the program at every time is not the same. Even if the node selected at the first time is the same, the node selected at the second time or the third time is different. The problem indicates the algorithm is not stable. In this paper, we make an improvement to let the source node selected in each time is fixed, and then the algorithm has a certain stability.

Based on the improvement, we selected the quality function of local community detection algorithm in [16] which first put forward the concept of the local community detection. The quality function R is:

\[ R = \frac{I}{T} = \frac{B_{\text{in}}}{B_{\text{in}} + B_{\text{out}}} \]

Where the number of edges is indicated by T, the number of edges end points is in B., I is the number of those edges with only one end point in B(B is the boundary of the community C) and the
other end point in C (C is the extracted community). $B_{in}$ is the same as I, $B_{out}$ is the number of edges with only one end point in B and the other end point in N (N is the neighbors of the community C).

The procedure of our method is as the following:
(i) pick a node A with the maximum centrality;
(ii) the natural community of node A is detected;
(iii) a node B is picked with the maximum centrality and it is not assigned to any group;
(iv) the natural community of B is detected, and is explored all nodes regardless of their possible membership to other groups;
(v) repeat from 3.

As to the question about why we select the node with the maximum centrality at every time, it because that we consider the node itself has a big possibility to be an overlapping node. So on the practical significance, the community detected based on the source node like this has a big possibility to be an overlapping community.

4. Experimental Results and Analysis
In our experiments, we compare the normalized mutual information (NMI) of [14] and our method. We perform experiments with two real-world networks---- the NCAA Football network and the Zachary’s karate club.

4.1 NCAA Football Network
The dataset contain 179 vertices (60 outliers) and 787 edges. See Fig.3.

![Figure 3](image1.png)

**Figure 3.** NCAA Football network. The boundary vertices which only have 1 degree are outliers. The vertices are in the same community with the same color.

![Figure 4](image2.png)

**Figure 4.** The detection of Fig.3 by [14]. overlapping nodes are the black nodes.
Figure 5. The detection of Fig.3 by our method. The overlapping nodes are the black.

From the experiment, the value of NMI in our method is 0.38 while in [14] is 0.42. The result of our method is slightly lower than [14]. And the result of running the program at every time in our method is the same while in [14] is not.

4.2 Zachary’s Karate Club
The dataset contain 34 vertices and 78 edges. We use the data set adopted the best cover put forward in [14] is divided into two communities, having five overlapping nodes (3,9,10,14 and 31). See the Fig.6.

Figure 6. Zachary’s karate club. the overlapping nodes are green nodes.

Figure 7. The detection of Fig.6 by [14]. The black nodes are overlapping.
Figure 8. The detection of Fig.6 by our method. The green nodes are overlapping nodes.

The value of NMI is 0.34 (our method) and 0.31 ([14]) and the result of our method is slightly higher than [14].

5. Conclusion
We have reviewed the algorithm in [14], and then we improved the algorithm. Our improvement makes the algorithm has a certain stability rather than random. The experiments show that our results are similar with [14]. Sometimes [14] is better than our method, but sometimes [14] is also worse than our method.

6. References
[1] S M 1969 Physics of Semiconductor Devices (New York: Wiley–Interscience).
[2] Albert, R. & Barabasi, A.-L. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97 (2002).
[3] Barabasi, A.-L. & Albert, R. Emergence of scaling in random networks. Science 286, 509–512 (1999).
[4] Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998).
[5] Mendes, J. F. F. & Dorogovtsev, S. N. Evolution of Networks: From Biological Nets to the Internet and WWW (Oxford University Press, Oxford, 2003).
[6] Danon L, Duch J, Arenas A and D’iaz-Guilera A 2007 Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science, eds. Caldarelli G and Vespignani A (World Scientific, Singapore) 93
[7] Girvan M and Newman M E J 2002 Proc. Natl. Acad. Sci. 99 7821 Newman M E J 2004 Eur. Phys. J. B 38 321
[8] Fortunato S and Castellano C 2009 in Encyclopedia of Complexity and System Science, ed. Meyers B (Springer, Heidelberg), arXiv:0712.2716 at www.arXiv.org
[9] Gregory, S.: An Algorithm to Find Overlapping Community Structure in Networks. In:Kok, J.N., Koronacki, J., Lópe de Mántaras, R., Matwin, S., Mladenıš, D., Skowron, A.(eds.) PKDD 2007. LNCS (LNAI), vol. 4702, pp. 91–102. Springer, Heidelberg (2007)
[10] Baumes, J., Goldberg, M., Magdon-Ismail, M.: Efficient Identification of Overlapping Communities. In: Kantor, P., Muresan, G., Roberts, F., Zeng, D.D., Wang, F.-Y., Chen, H., Merkle, R.C. (eds.) ISI 2005. LNCS, vol. 3495, pp. 27–36. Springer, Heidelberg (2005)
[11] Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. Nature 435, 814–818 (2005)
[12] Zhang, S., Wang, R., Zhang, X.: Identification of Overlapping Community Structure in Complex Networks Using Fuzzy C-means Clustering. Physica A: Statistical Mechanics and its Applications 374(1), 483–490 (2007)
[13] Gregory S. A Fast Algorithm to Find Overlapping Communities in Networks[J]. Lecture Notes in Computer Science, 2008:408-423.
[14] Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks[J]. New Journal of Physics, 2009, (15).
[15] Newman M. Detecting community structure in networks. The European Physical Journal B-Condensed Matter and Complex Systems, 2004, 38(2): 321-330.

[16] Clauset A. Finding local community structure in networks[J]. Physical review E, 2005, 72(2): 026132.

[17] Zhengyou Xia, Z. Bu, Community detection based on a semantic network, Knowledge Based System, Volume 26, February 2012, Pages 30–39.