Community Detection Research Based on Line Graph

Chun Gui1,2,*

1Key Laboratory of China’s Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou City, Gansu Province, 730000 China
2College of mathematics and computer science, Northwest Minzu University, Lanzhou City, Gansu Province, 730000 China

*Corresponding author e-mail: 281071544@xbmu.edu.cn

Abstract. Community structure is an important property of complex networks, in which network nodes are tightly split into different groups. In this paper, a new overlapping community detection algorithm based on line graph is proposed. Spectral analysis based on Normal Laplace Matrix of line graph is introduced. We not only make the spectral clustering to be an overlapping algorithm but also define a new similarity between edges by spectral analysis. We test our method on three-community network and Zachary karate club network. The experimental result shows that our method can detect community structure and overlapping nodes effectively. It is worth to note that our method is against undirected and unweighted graphs.

Keywords: Complex Network, Normal Laplace Matrix, Spectral Clustering, Overlapping Community

1. Introduction

Community detection is a crucial field in complex network study. Community detection in complex network helps in understanding the network topology and dynamic behavior of the network. However, in real complex networks, the nodes often belong to more than one community, which leads to the overlap between a community and another community [1]. For instance, in protein-protein interaction networks, many proteins belong to more than one functional community and a node might have multiple functions, namely protein complexity [2]. In the past several years, many researchers have put efforts into overlapping community detection and proposed many algorithms. For example, Palla first proposed detecting overlapping community algorithm CPM (Clique Percolation Method) in 2005 [3]. CPM is suitable for networks with dense connected parts [4]. Lancichinetti proposed overlapping algorithm based on Local Expansion and Optimization [5].

Gregory proposed fuzzy community detection algorithms [6]. A defect of such algorithms is that they have to determine the dimensionality k of the membership vector. Remarkably, Evans explored the idea of line graph to discover community structure [7], which defines community as a partition of the links rather than the set of nodes. A comparison of different community overlapping algorithms can be found in a survey given by JIERUI XIE [4].
In this paper, we select three-community network (including 19 nodes and 37 edges) and Zachary karate club network to test our method. The experiment result shows us that our method can effectively and correctly detect community structure compared with the previous results. The rest of the paper is organized as follows. Section 2 proposes our overlapping community detecting algorithm. Section 3 details the experiment and result. Section 4 concludes our work.

2. Proposed Method

We employ spectral analysis based on Normal Laplace Matrix [8] to define the similarity of elements. First, it needs to seek the Normal Laplace Matrix of line graph. Normal Laplace Matrix is defined as $L = D^{-1}A$, where D is the degree matrix and A is the adjacency matrix of the graph. In order to obtain the Normal Laplace Matrix of a line graph, the first work is calculating the degree matrix of line graph. In a line graph, define the elements of the line graph as $e_{ij}$, the degree of $e_{ij}$ is the number of other elements which connect with $e_{ij}$ ($v_i$ and $v_j$ are the two vertex of edge $e_{ij}$). According to the relationship between line graph and original network, we can get formula as follows:

$$\text{deg}(E_{ij}) = \text{deg}(v_i) + \text{deg}(v_j) - 2$$

According to graph theory, the degree of the vertex is equal to the sum of corresponding row (or column) elements in adjacency matrix. We define the adjacency matrix of line graph as $A(L(G))$, and $A(L(G)) = (e_{mn})_{q \times q}$ (q is the number of edge in original network), so the degree of elements $e$ as follows:

$$\text{deg}(e_m) = \sum_n e_{mn}$$

According to the definition of degree matrix, get the degree matrix $D(L(G))$ of a line graph $L(G)$ as follows:

$$D(L(G)) = \begin{bmatrix} \text{deg}(e_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \text{deg}(e_q) \end{bmatrix}$$

Afterwards, according to the definition of Normal Laplace Matrix, we can get the Normal Laplace Matrix $N(L(G))$ of line graph as follows:

$$N(L(G)) = D(L(G))^{-1}A(L(G))$$

With the Normal Laplace Matrix $N(L(G))$ of line graph, the next work is calculating the eigenvalues and eigenvectors of Normal Laplace Matrix. The maximum eigenvalues of Normal Laplace Matrix is always equal to 1, which is called trivial eigenvalue. According to the principle of spectral analysis based on Normal Laplace Matrix proposed by Capocci [8], for a network with apparent community structure, assuming the number of community is $m$, the Normal Laplace Matrix has m-1 nontrivial eigenvalues close to 1, corresponding to the m-1 nontrivial eigenvectors. For these m-1 nontrivial eigenvectors, the components corresponding to nodes belong to the same community are very approximate and the profile of each eigenvector appears step-like. A step represents a community and the number of the step represents the number $m$ of community in the network. The advantage of the idea is that it doesn’t need priori knowledge of the number of community. Taking advantage of Capocci’s method [8] define the first nontrivial eigenvector as the similarity in line graph. Similarly, the profile of the first eigenvector of Normal Laplace Matrix $N(L(G))$ of line graph appears step-like. The details of experimental proof procedure will be presented in Section 3. Once that profile appears step-like, we can obtain the community structure of line graph easily, and one step is a community. Finally, the nodes in the line graph are recovered to the links in the original network. Thus, we can detect overlapping nodes because a link often exists for one dominant.
To evaluate the clustering results, we use partition density $D$ to measure the clustering result. For a network with $M$ links, $\{p_1, p_2, \ldots, p_c\}$ is a partition of the links into $C$ subsets. Subset $p_c$ has $m_c = |p_c|$ links and $n_c$ nodes. The formula of partition density $D$ is as follows:

$$D = \frac{\sum \frac{m_c}{\sum m_c} \sum \frac{m_c(n_c-1)}{(n_c-2)(n_c-1)}}{\sum m_c}$$

(5)

Different $k$ values corresponding to different $D$, compare the partition density $D$ for different $k$ values and pick out the best community partitioning.

The basic steps of the algorithm are described as follows:

**STEP 1:** Read the network and obtain the incident matrix $B(L(G))$.

**STEP 2:** According to formula $A(L(G)) = B^TB - 2I_q$ ($I_q$ is $q$ order unit matrix), obtain the adjacency matrix $A(L(G))$ of line graph.

**STEP 3:** According to the definition of the degree of line graph proposed in this section, calculate the sum of corresponding row (column) elements in adjacency matrix $A(L(G))$ and obtain the degree matrix $D(L(G))$ of line graph.

**STEP 4:** According to formula (5), calculate the Normal Laplace Matrix $N(L(G))$.

**STEP 5:** Calculate the eigenvalues and eigenvectors of Normal Laplace Matrix $N(L(G))$.

**STEP 6:** Select the first nontrivial eigenvectors and obtain the distribution of the first nontrivial eigenvectors. If the distribution appears step-like, define the first nontrivial eigenvector as the similarity and complete the spectral analysis. Otherwise, go to **STEP 7**.

**STEP 7:** Set the number of community $k$, select the first nontrivial eigenvectors as clustering sample of $k$-means algorithm, calculate the partition density $D$ for different $k$, respectively, and pick out the best partition.

**STEP 8:** Analysis the clustering result of **STEP 6** or **STEP 7** and find out overlapping communities.

### 3. Experiments and Results

#### 3.1. Three-Community Network

The network in Fig. 1 has 19 nodes and 37 edges, three clear clusters appear, composed by nodes 0–6, 7–12 and 13–19. The distribution of the first nontrivial eigenvector components for Normal Laplace Matrix is shown in Fig. 2. The community structure is presented in Table 1. From Table 1, we find overlapping node between community A and community B is node 6, B and C is node 7, C and D is node 10, D and E is node 13.
Table 1. The clustering result of line graph

| Community | Set               |
|-----------|-------------------|
| A         | {1-0,2-0,2-1,3-1,4-1,3-2,6-2,4-3,6-3,5-4,6-4,6-5} |
| B         | {6-7}             |
| C         | {8-7,9-7,10-7,12-7,9-8,10-8,11-8,12-8,10-9,12-9,11-10,12-11} |
| D         | {13-10}           |
| E         | {14-13,16-13,18-13,15-14,16-14,17-14,16-15,18-15,17-16,18-16,18-17} |

Fig.2 is the distribution of the first nontrivial eigenvector of the line graph of Three-Community Network. However, in Capocci’s article [8], the author only presented three communities structure to us, while did not detect the overlapping or important nodes. Therefore, our method is effective to detect community structure and overlapping nodes.

3.2. ZACHARY--KARATE CLUB

The ZACHARY--KARATE CLUB dataset is collected from the members of a university karate club by Wayne Zachary in 1997[9]. According to our method, first we calculate the eigenvalues and eigenvectors of Normal Laplace Matrix of line graph. The distribution of first nontrivial eigenvector components is presented in Fig.3, whose eigenvector profile is not step-like. Thus, we can’t obtain the community structure and the number of community intuitively. Due to the limitation, we set the number $k$ of community in advance. We set $k = 2,3,4,5,6$ and calculate the partition density $D$, respectively. The result is presented in Table 2. We can find that the partition density $D$ is maximum corresponding to the $k = 2$. Thus, we consider that the network has two communities.
The clustering result of method was shown in Fig.4. Compared with the result in Clara’s article [10], the identified overlapping nodes 3, 9, 10, 14 and 32, our result additionally identified overlapping nodes 20, 29 and 31.

4. Conclusion
In this paper, we defined a new similarity between edges by spectral analysis and make spectral clustering algorithm based on Normal Laplace Matrix to be an overlapping community detecting algorithm. We applied spectral clustering algorithms to line graph and have demonstrated a variety of applications of our methods to real-world networks. The clustering result of our method is in accord with the actual situation. Therefore, our method could be useful and effective in detecting overlapping community.
Table 2. The clustering result of line graph Different $k$ corresponding to the partition density $D$

| the number of community($k$) | of the partition density($D$) |
|------------------------------|-----------------------------|
| 2                            | 0.1                         |
| 3                            | 0.08041                     |
| 4                            | 0.07041                     |
| 5                            | 0.05196                     |
| 6                            | 0.06428                     |

Acknowledgments
This work is partly supported by the Fundamental Research Funds for the Central Universities 2019 (No.31920190029). Gansu Higher Education Innovative Ability Improvement Project(No.2020B-068). The education and teaching reform project (No.2019XJJG-56). Talent introduction project (No.Z20051)

References
[1] Yong Yeol Ahn, James P, Bagrow, Sune Lehmann. Link communities reveal multiscale complexity in networks [J]. Nature, 2010, 466(7307):761-764.
[2] K. M. Hall. An r-dimensional Quadratic Placement Algorithm [J]. Management Science 17, 1970, 219-229.
[3] PALLA G., DERENYI I, FARKAS I, et al. Uncovering the overlapping community structure of complex networks in nature and society [J]. Nature, 2005, 435: 814-818.
[4] Jierui Xie, Stephen Kelley, Boleslaw K.Szymanski. Overlapping community detect-ion in networks: the state of the art and comparative study. ACM Computing Surveys, 2012, 45(4).
[5] Juan, Ignacio, Perotti. Hierarchical mutual information for the comparison of hierarchical community structures in complex networks [J]. Physical Review E, 2015.
[6] Gregory, S. Fuzzy overlapping communities in networks [J]. J. Stat. Mech. 2011, 02, P02017.
[7] Ding Z, Zhang X, Sun D, et al. Overlapping Community Detection based on Network Decomposition [J]. Rep, 2016, 6:24115.
[8] Capocci A, Servedio V D P, Caldarelli G, Colaio ro F. Detecting communities in large networks [J]. Computer Science, 2004 (3243):181-187.
[9] Zachary W W. An information flow model for conflict and fission in small groups [J]. Journal of anthropological research, 1977:452-473.
[10] Clara Pizzuti. GA-NET: a genetic algorithm for community detection in social networks. In Proceedings of the 10th International Conference on Parallel Problem Solving from Nature (PPSN),2008, 1081–1090.