Overlapping Community Detection Combining Topological Potential and Trust Value of Nodes

Xiaohong Li, Weiying Kong, Weiyi Wei, Enli Fu, Huifang Ma

Overlapping Community Detection Combining Topological Potential and Trust Value of Nodes

Xiaohong Li[1]*, Weiying Kong[1], Weiyi Wei[1], Enli Fu[2], Huifang Ma[1]

1. College of Computer Science and Engineering, Northwest Normal University, Lanzhou China
xiaohongli@nwnu.edu.cn
2. School of information science and engineering, Lanzhou University, Lanzhou China
2384371518@qq.com

Abstract. Aiming at the problems of existing algorithms, such as instability, neglecting interaction between nodes and neglecting attributes of node, an overlapping community discovery algorithm combining topological potential and trust value of nodes was proposed. Firstly, the importance of nodes is calculated according to topological potential and the trust value of the node, and then K core nodes are selected. In final, the final division of communities are finished by using the extended modularity and core nodes. Experimental results on LFR network datasets and three real network datasets, verify the efficiency of the proposed *OCDTT* algorithm.

Keywords: Overlapping Community Detection, Importance of Node, Topological Potential, Trust Value.

1 Introduction

A complex network is an abstraction of a complex system. Most real-world networks, such as transportation, social or gene-regulatory networks, are complex networks. The "community" refers to the set of nodes with the same or similar characteristics in a complex network[1],there are overlapping and non-overlapping communities in complex network, which reflects the different agglomeration of nodes.

So far, there have been many research algorithms for the study of overlapping community research. Li[2] detect overlapping communities in the unweighted and weighted networks with considerable accuracy. Gregory[3] propose overlapping community detection algorithm based on nodes splitting, according to betweenness and edge-betweenness. Zhang[4] is inspired by label propagation and modular optimization, they introduce a community detection algorithm based on fuzzy membership propagation. In each iteration, candidate seed of potential communities are selected using topological features, and then the membership of the selected seeds is propagated to non-seed vertices, thus multiple communities could be obtained. Ahn[5] propose LINK algorithm performing hierarchical clustering on links based on idea of transforming in overlapping and non-overlapping communities in link networks. Although, researchers achieve great achievements in the research of community detection[6,7]. Finding communities is a very challenging and promising research field. how to distinguish stable overlapping communities and propose efficient algorithms is still one of hot problems for many researchers.

We propose an overlapping community discover algorithm combining the topological potential and the trust value of nodes (Abbreviated as *OCDTT*) against problems appeared in existing algorithms. Firstly, the importance of nodes $I(v_i)$ in the network is computed combining the topological potential of nodes and the trust value of the node, and then K core nodes are selected according to $I(v_i)$. Based on the core nodes, communities are extended using the extended modularity.

2 Preliminary Study

We review the existing concepts, and define basic concepts and the problem of community detection. Table 1 gives a list of symbols used in this paper.

Table 1. Symbols

Symbols	Definition
$\Gamma(v_i)$	*the set of p-order neighbor nodes of* v_i.
$s(v_i, v_j)$	*the similarity between node* v_i *and* v_j
c	*The size of the set* $\Gamma(v_i)$
α	*the balance factor*
$\varphi(v_i)$	*the topological potential of node* v_i
$t(v_i)$	*The trust value of node* v_i
$I(v_i)$	*The importance of node* v_i
λ	*regulation parameter*$(0< \lambda <1)$

2.1 Trust Value of Nodes

***Definition*1** If $path_{min}(v_i, v_j) = p$, the node v_j is called as the p-order neighbor node of node v_i.

For example, when $path_{min}(v_i, v_j) = 1$, v_j is 1- order neighbor node of v_i. The trust value of the node v_i is defined as the sum of the similarity $s(v_i, v_j)$ between v_i and all its p-order neighbors[8]. $\boldsymbol{v}_i^A = (a_i^1 \cdots, a_i^d)$ is attribute vector of node v_i in formula (1).

$$t(v_i) = \sum_{v_j \in \Gamma(v_i)} s(v_i, v_j) \qquad \textit{and} \qquad s(v_i, v_j) = \frac{\boldsymbol{v}_i^A \cdot \boldsymbol{v}_j^A}{\|\boldsymbol{v}_i^A\| \times \|\boldsymbol{v}_j^A\|} \tag{1}$$

2.2 Topological Potential of Nodes

Potential refers the work generated by a particle moving from one point to another point in the field, and the work may depend only on the position of the particle and not depend on the path along which the particle moved in the physics science. If each node in a complex network is regarded as a particle in the field, and the edge between the nodes is used as a link between the particles to generate work[9]. Then the concept of potential field is applied to complex networks, which makes the connections between nodes have fine physical features and stability. Considering that the work will decrease with the increase of the shortest path length between nodes, So the topological potential of node v_i in the complex network G can be improved as formula (1). Impact factor $\delta \in (0, +\infty)$ is used to control impact scope of each node.

$$\varphi(v_i) = \sum_{k=1}^{p} count(k) * e^{-(\frac{path_{min}(v_i, v_j)}{\delta})^2} \tag{2}$$

Here, *count*(k) is the number of the k-th order neighbors of node v_i. Keeping k constant, $\varphi(v_i)$ get larger as *count*(k) gets larger, the network is densely. Therefore, the topological potential in equation (2) reflects the intensity of the network.

3 Implementation on Our Framework

In this research, we aim to solve community detection problem and describes how they work under this framework, the overall procedure our algorithm is shown in Fig. 2.

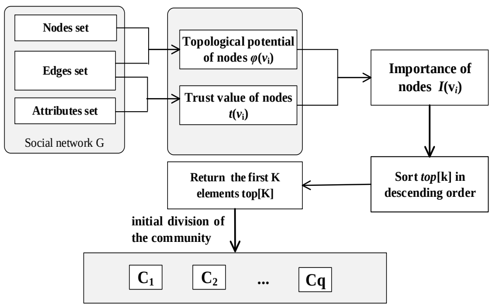


Fig. 1. Framework of OCDTT Algorithm

3.1 Select Core Nodes

Due to the different positions of nodes and the difference of the interaction between nodes in complex networks, which make each node has different importance, their contributions to the network are also different[10]. Thus we can estimate the importance of nodes $I(v_i)$ from the structure of network and attributes of node.

$$I(v_i) = \alpha\varphi(v_i) + (1-\alpha)t(v_i) \tag{3}$$

Next, top-k nodes are selected as core nodes according to the value of $I(v_i)$.

```
Input: Network G =(V, E, A),σ,α,K
Output: Core node set top[]
1. initialize top[]={0}, k=1;
2. for each vi∈V do
```

2.1 for each $v_j \in V$ and $j \neq i$ do
  if $1 < path_{\min}(v_i, v_j) < p$ then
    2.1.1 compute topological field of node v_i;
    2.1.2 compute trust value of the node v_i;
2.2 get the value of $I(v_i)$ by formula (5)
2.3 $top[k] = I(v_i)$;
3. Sort the array top[k] in descending order, and select the first K elements to output;
4. return $top[K]$;

3.2 Division of The Communities

We selects adjacent node of the core nodes in turn, and tries to add it to a community and calculate the value of *EQ*[11].

$$EQ = \frac{1}{2m} \sum_{k} \sum_{v_i \in C_k, v_j \in C_k} \frac{1}{O_{v_i} O_{v_j}} [M_{ij} - \frac{d_{v_i} d_{v_j}}{2m}] \quad (4)$$

If adding of the adjacent node makes EQ increases, then add the node into the community. Otherwise, select other adjacent node and repeat such operations until all adjacent nodes of core nodes is retrieved. Procedure is as follows:

Input: $G=(V,E,A)$, core nodes set top[K]
Output: Initial community $C = \{C_1, C_2, \ldots, C_K\}$
1. $C = \Phi$;
2. Select nodes from core nodes to expand communities:
for each $v_i \in top[K]$, i=1,2, …K do
2.1 $C_i = \{v_i\}$
2.2 if($EQ(C_i \cup v_j) > EQ(IC_i)$) then $C_i = C_i \cup v_j$
2.3 $C = C \cup IC_i$
3. return C;

4 Experimental Results and Analysis

To compare and contrast the performance of the *OCDTT* method, we apply it to a variety of two datasets. We use Normalized Mutual Information(*NMI*) to evaluate performance of communities finding.

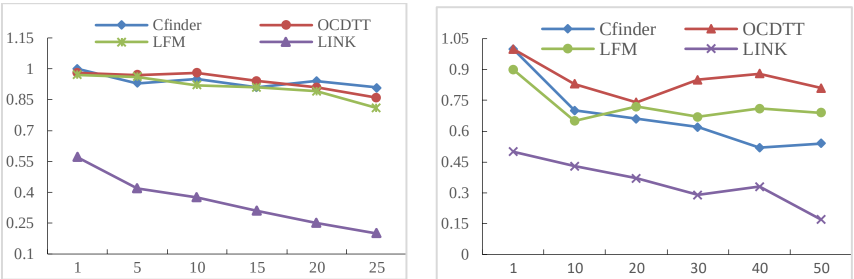
4.1 Experimental Results and Analysis

Complex simulation LFR network which internal structure and sizes of communities are scalable to demonstrate effectiveness of our method. Table 1 illustrates four groups parameter information of simulation LFR network.

Table 2. two groups of *LFR* network parameter information

Name	N	K_{avg}	K_{max}	C_{min}	C_{max}	μ	O_m	O_n
S_1	1000	15	50	10	25	0.1	2	0~20
S_2	1000	10	50	10	50	0.1	2	0~10

For LFM algorithm, range of parameter α is [0.8, 1.6] and step is 0.1. In CFinder, k is integer in [3, 8]. Range of threshold in *LINK* is [0.1, 0.9] and step is 0.1. Impact factor of *OCDTT* σ=1.034. Considering that structure and attributes have equal impact, set balance factor σ=0.5.



(a) NMI on S_1 (O_m=2) (d) NMI on S_2 (O_m=5)

Fig. 2. Results of four algorithms on different subsets

We can study by observing figure 2 that *OCDTT* is more outstanding to partition communities in most conditions. It indicates that more communities a node belongs to, more complex the network is, lead to poorer performance of algorithms. It can be observed through comparing results of S_1 and S_2 that *OCDTT* is more stable than others when the number of communities increases.

4.2 Results and Analysis on Real Datasets

Three real networks consist of Karate, Dolphins and Football. Karate is friend network of 34 members in a Karate club. The network has 78 edges representing relations among 34 members. Dolphins dataset is about dolphin groups in New Zealand. The network represents biological family relations of 62 dolphins having 159 edges. Football data represent groups of match teams in a university including 115 members and 613 relations. It can be seen from table 3 that *EQ* is not less than 0.5000 after partition in *OCDTT*. The partition performance is remarkable. *LINK* algorithm produces linking communities of small sizes when implementing. It hampers form of communities, so corresponding result is the worst.

Table 3. EQ value on real network

Name	LFM	CFinder	LINK	OCDTT
Karate	0.3740	0.1860	0.0270	0.3720
Dolphins	0.4360	0.3610	0.1490	0.5000
Football	0.5260	0.4880	0.0830	0.5650

5 Conclusion

We combine the topological potential of nodes and the trust value of nodes to compute importance of nodes in the network. Start from those selected core nodes, utilize expanding module function to generate final community partition. Experimental results demonstrate that *OCDTT* could achieve better results comparing to algorithms of the same kind. In further work, we will continue partitioning communities in aspects of structure and attributes. Based on this, optimize *OCDTT*. Strengthen efficiency of the implement. Attempt to apply the algorithm to real network analysis and web recommendations.

6 Acknowledgments

This work was supported in part by National Natural Science Foundation of China (No. 61762078, 61862058, 61967013), Youth Teacher Scientific Capability Promoting Project of NWNU (No. NWNU-LKQN-16-20).

References

1. Duch J, Arenas A. Community detection in complex networks using extremal optimization. J. Physical review E, 2005, 72(2): 027104
2. Li J, Wang X, Cui Y. Uncovering the overlapping community structure of complex networks by maximal cliques. J. Physica A: Statistical Mechanics and its Applications, 2014, 415: 398-406
3. Gregory, Steve. An algorithm to find overlapping community structure in networks. In: European Conference on Principles of Data Mining and Knowledge Discovery. Springer, Berlin, Heidelberg, 2007 : 91-102
4. Zhang H, Chen X, Li J, et al. Fuzzy community detection via modularity guided membership-degree propagation. J. Pattern Recognition Letters, 2016, 70: 66-72
5. Ahn Y Y, Bagrow J P, Lehmann S. Link communities reveal multiscale complexity in networks. J. nature, 2010, 466(7307): 761-764
6. Gui Q, Deng R, Cheng X, et al. A New Method for Overlapping Community Detection based on Complete Subgraph and Label Propagation. In: 3rd International Conference on Intelligent Information Processing. 2018: 127-134.
7. ZhiXiao W, Zechao L, Xiaofang D, et al. Overlapping community detection based on node location analysis. J. Knowledge-Based Systems, 2016, 105: 225-235
8. Wang Z, Zhao Y, Chen Z, et al. An improved topology-potential-based community detection algorithm for complex network. J. The Scientific World Journal, 2014
9. Shen H , Cheng X , Cai K , et al. Detect overlapping and hierarchical community structure in networks. J. Physica A, 2009, 388(8):1706-1712
10. Bin W, Yayun W, Jinfang S, Zejun S. Identifying Influential Nodes of Complex Networks Based on Trust value. J. Journal of Chinese Computer Systems, 2019, 40(11): 2337-2342
11. Moayedikia A. Multi-objective community detection algorithm with node importance analysis in attributed networks. J. Applied Soft Computing, 2018, 67: 434-451