TELKOMNIKA, Vol. 11, No. 12, December 2013, pp. 7649~7654
e-ISSN: 2087-278X

Community Structure Detection Algorithm based on the Node Belonging Degree

Jian Li*, Huiwen Deng
School of Computer and Information Science, Southwest University, Chongqing, P. R of China, 400715
Corresponding author, e-mail: lijiansym@qq.com, huiwend@swu.edu.cn

Abstract

In this paper, we propose a novel algorithm to identify communities in complex networks based on the node belonging degree. First, we give the concept of the node belonging degree, and then determine whether a node belongs to a community or not according to the belonging degree of the node with respect to the community. The experiment results of three real-world networks: a network with three communities with 19 nodes, Zachary Karate Club and network of American college football teams show that the proposed algorithm has satisfactory community structure detection.

Keywords*: complex network, community structure, clustering coefficient, node belonging degree*

1. Introduction

As one of the characteristics of the complex network in physical meaning and mathematical features in-depth study, it was found that many of the actual networks have the community structure [1]. The entire network consists of a number of "groups" or "clusters", in which network nodes are joined together in tightly knit groups, among them there are looser connections.

So far, people have proposed many algorithms to find the community structure of complex networks. Kernighan-Lin algorithm [2] divides the network into two communities which sizes are known based on greedy algorithm. Wu and Huberman present a method [3] which is based on notions of voltage drops across networks that are both intuitive and easy to solve regardless of the complexity of the graph involved. GN algorithm [4] works by using information about edge betweenness to detect community peripheries. Newman and Girvan proposed a different technique [5], which focuses on currents flowing on edges of a network in order to discover communities. Clique percolation [6], introduced by Palla, Derényi and Vicsek Nature, is a deterministic community detection method which allows for overlapping communities and is purely based on local topological properties of a network. On the basis of the above algorithm, Kumpula and others present a sequential clique percolation algorithm (SCP)[7] to do fast community detection in weighted and unweighted networks. Another view is that the problem of overlapping community detection can be converted into traditional fuzzy clustering problem, through the calculation of fuzzy membership degree to reveal the relationship of nodes and communities [8]. LMF algorithm [9], which consists of finding the local maxima of a fitness function by local, iterative searching, enables each node to be included in more than one module, leading to a natural description of overlapping communities. Mixture models of community structure were built by the use of probability method to make a study of community structure of complex networks to find the largest community structure [10]. Evans and Ahn see the edge as the research object to identify communities in complex networks respectively [11, 12].

Although the algorithms above can solve the problem of detecting the communities, but the complexity of those are all associated with the structure of the entire network and need higher time and space complexity. In this paper, according to the belonging degree of a node with respect to the community, we present a different method to determine whether the node belong to the known community or not. This method only according to the network's local features discovers the community structure of the network, ignoring the size of the network. In this paper, the node belonging degree defined is different from membership degree [8].

Received June 30, 2013; Revised August 19, 2013; Accepted September 4, 2013

2. Basic Concepts and Algorithm Description

2.1. Node Clustering Coefficient

In this paper, the network is simple undirected graph without isolated nodes.

Suppose the degree of node *v* is *k* and the number of the edges between the neighbor nodes of node *v* is *e* in the network *G* = (*V*, *E*). The clustering coefficient of node *v* is defined by:

$$c = \frac{2e}{k(k-1)} \tag{1}$$

Obviously, the smaller clustering coefficient of one node says the neighbor nodes of this node have loose connection with each other. On the other hand, the larger clustering coefficient of one node indicates the tightly connections between the neighbor nodes, the more likely the neighbor nodes belong to the same community.

2.2. The Node Belonging Degree

2.2.1. Definition of the Node Belonging Degree

Definition: Suppose *C* is a community in the network *G* = (*V*, *E*), which is a subgraph induced by some nodes of *G*. For node *v* in *G*, the belonging degree of node *v* with respect to community *C* is defined as:

$$b = \frac{m}{k} \tag{2}$$

Where *k* is the degree of node *v*, *m* is the number of neighbor nodes of node *v* in community *C*.

We know from the definition that the belonging degree of node with respect to the community ranges from 0 to 1. The closer to 1 the value *b*, the tighter the node *v* belongs to the community; and the closer to 0 the looser the node *v* belongs to the community. We merge the node into community based on the node belonging degree.

2.2.2. Threshold of Node Belonging Degree

We give a threshold b_0 of node belonging degree. If the belonging degree of a node with respect to a community is greater than the given threshold b_0, we incorporate the node into this community. However, if the given threshold b_0 is too small, there may be multiple nodes belonging to more than one community and this will affect the division of community. Setting the threshold of belonging degree $b_0 \geq 0.5$ can solve this problem, because the belonging degree of one node with respect to multiple communities cannot be greater than 0.5 at the same time. Hence, we set the threshold of belonging degree $b_0 = 0.5$. If too many nodes whose node belonging degree with respect to the community is less than b_0, we can reduce the threshold b_0. (In the further work, we want to consider how to appropriately reduce *b*, temporarily not considered in this paper).

It is possible that there exists a node whose belonging degree with respect to a known community is less than b_0, but if we add this node into the known community, the belonging degree of its neighbor node with respect to the known community will greater than b_0 and the belonging degree of this node with respect to the known community will also become greater than b_0 after adding its neighbor node into the known community. As shown in Figure 1, supposing we set the belonging degree threshold $b_0 = 0.5$, either the belonging degree of node 1 with respect to the known community or belonging degree of node 2 with respect to the known community is 0.25, and then they cannot be added into the known community at this moment. But if we put one of them into the known community, all of the belonging degree of them become 0.5. As a result, they can be added into the known community.

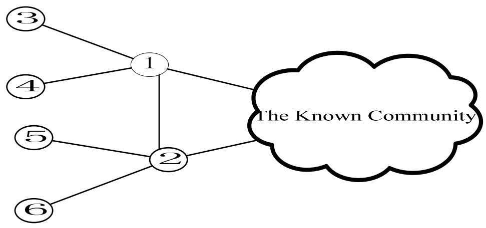

Figure 1. The Discussion of Belonging Degree Threshold

Therefore, by assuming a smaller belonging degree threshold b_1 we detect the initial community roughly, and then we use a larger belonging degree threshold b_2 to verify whether the belonging degree about every nodes with respect to the initial community is greater than b_2 or not. If the node cannot meet our requirement, remove it.

2.3. Description of the Algorithm

The outside nodes of a community which directly connect with a node inside the community are defined as neighbor nodes of the community.

We define the two belonging degree thresholds are b_1 and b_2, and $G_j = (V_j, E_j)$ is the j-th community in network $G = (V, E)$. First, by computing the clustering coefficients between a node and its neighbor nodes whose degree is greater than 2, we can find the largest one v_i, set node v_i and its neighbor nodes as the initial community G_j. Next, we compute the belonging degree b of nodes which are neighbor nodes of community G_j and add the nodes whose belonging degree $b > b_1$ into the community G_j. Finally, we calculate the belonging degree of nodes in community G_j, and remove the nodes whose belonging degree of community G_j with $b < b_2$. Repeat the whole process above until all of the communities are found. There may be some nodes in the network whose belonging degree of every known community is smaller than b_2, adding each of them into the community where the belonging degree of the node with respect to this community is the maximal respectively.

We describe the detail of the algorithm as follows.

INPUT: one undirected network $G = (V, E)$, belonging degree thresholds b_1, $b_2(b_2 > b_1)$, and set $V_0 = V$, $j = 0$.

OUTPUT: the community structure of network G.

Step 1. If there aren't nodes in V_0 whose degree greater than 2 computed in G, do **step 5**. Otherwise set $j = j + 1$.

Step 2. Select a node whose degree greater than 2 in V_0, compute all the clustering coefficients of this node and its neighbor nodes in V_0 whose degree greater than 2. Choose the largest clustering coefficient node, set this node and its neighbor nodes in V_0 as V_j, and get the initial community G_j.

Step 3. Compute the belonging degree b of every neighbor node of G_j in V_0 with respect to G_j, add the node into initial community G_j if $b > b_1$.

Step 4. Compute the belonging degree b between the nodes in V_j with respect to G_j, remove the nodes with $b < b_2$. Set $V_0 = V_0 - V_j$, return to **step 1**.

Step 5. If $j = 0$, return G, stop. If $j \neq 0$, we compute the belonging degree between the node in V_0 with respect to every community as above respectively, and then we add the node into the community which the belonging degree between the node with respect to the community is the greatest until V_0 becomes empty. Finally, we get the communities G_1, G_2, …, G_j, stop.

If there isn't node whose degree greater than 2 in network, we think the community structure of this network is not obvious and then we see the entire network as a community.

3. Experiment and Analysis

3.1. A Network with Three Communities with 19 Nodes

A network with three communities composed as shown in Figure 2, the number of nodes in this network is 19. The network can be clearly divided into three groups obviously [13].

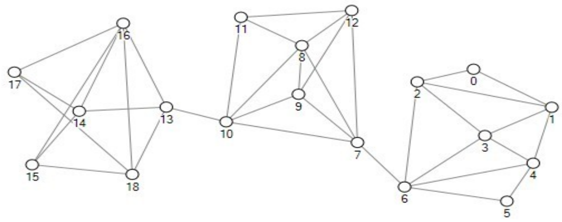

Figure 2. A Network with Three Communities with 19 Nodes

According to algorithm presented in this paper we set two belonging degree thresholds b_1 = 0.4, b_2 = 0.5. Select a node whose degree is greater than 2, node 1 for example. Compute the clustering coefficients of node 1 and its neighbor nodes whose degree greater than 2 and draw the clustering coefficient 2/3 of node 3 is biggest. So we set node 3 and its neighbor nodes as community $G_1 = (V_1,E_1)$ including node 3, 1, 2, 4, 6. Compute the belonging degree b of every neighbor node of G_1 with respect to community G_1 respectively as shown below. the belonging degree of node 0 with respect to G_1 is 1 which greater than belonging degree thresholds b_1, add node 0 into community G_1; the belonging degree of node 5 is also 1, and then we add node 5 into community G_1; the belonging degree of node 7 is 0.2 which is smaller than b_1, then we not need add the node 7 into G_1. As result, there isn't node whose belonging degree of G_1 is greater than b_1, therefore we obtain V_1 = {0,1,2,3,4,5,6}. Due to the limitation of space, this paper list only the process of getting the first community. Repeat the above steps will get three communities, the result is the same as in [13].

Experiments indicated that belonging degree thresholds b_1, b_2 in this network can be set to other values. If b_1 is in (0.25, 0.5) and b_2 is in (0.25, 0.75), the results are the same.

3.2. Zachary Karate Club

Zachary Karate club network [2] is a classical social network. In the early 1970s, Zachary constructed a relationship network about the members of karate club in a university. 34 nodes and 78 edges in the network, the node represents a club member, and the edge was the relationship between members. This network has been divided into two small communities and represented by the nodes 1 and 33. So, the algorithm is slightly modified as follows.

Set b_1 = 0.4 and b_2 = 0.5 as two belonging degree thresholds. Set node 33 and its neighbor nodes in community $G_1 = (V_1,E_1)$. Select a neighbor node of G_1, node 29 for example. Compute the belonging degree of node 29 with respect to G_1, get the belonging degree is 1 greater than b_1, so add the node 29 into the community G_1. And compute the belonging degree of other neighbor nodes with respect to community G_1, if the belonging degree greater than b_1, add the node into the community G_1 too. Meanwhile, computing the belonging degree of node in community G_1 with respect to G_1, if the degree less than b_2, remove the node from the community G_1. By this way, we get community G_1, and other nodes compose community G_2.

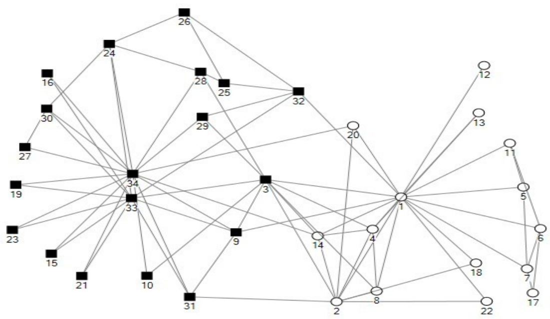

Figure 3. Zachary Karate Club Network

The community obtained by this algorithm is roughly the same with the result of the Kernighan-Lin algorithm[2]. As shown in Figure 3, it is consistent with [2] besides the node 3. In Figure 3, we can find the node 3 is between the two communities, the belonging degree between node 3 and each of the two communities is 0.5. The experiment result is consistent with [2] when the belonging degree thresholds b_1 in (1/3, 0.5) and b_2 in (0.5, 2/3).

3.3. College Football Network

College Football network is a network model constructed by Newman [5], it was about league football match between American college students in 2000. There were 115 nodes, 616 edges and 12 commuities in the network. Each node represents a football team, and each edge represents a match between two teams. This network is as shown in Figure 4.

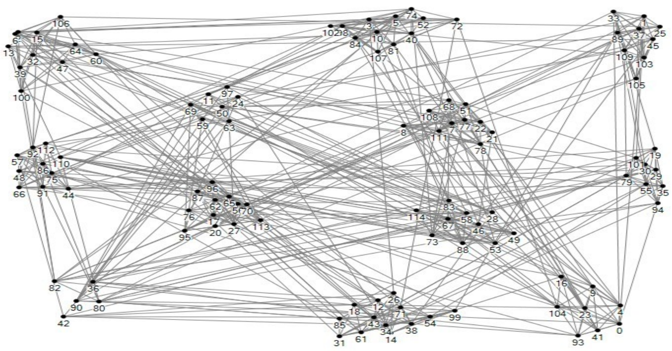

Figure 4. College Football Network

By algorithm, set b_1 = 0.3 and b_2 = 0.4. First select a node which degree greater than 2, for example node 0. Computing the clustering coefficient of node 0 and the neighbor nodes of node 0, we can find the largest clustering coefficient of the nodes are node 41 and node 93, they are all 0.53. Take one of them, such as node 41, add the node 41 and its neighbor nodes into community G_1. Compute the belonging degree between the community G_1 and its neighbor nodes, then compare the result with b_1, it is shown that they are all less than b_1. Computing the belonging degree of nodes in G_1 with respect to G_1, get the belonging degree of the node 8, 40, 67 is 0.2, 0.1, 0.25 respectively. They are all less than b_2, remove these three nodes from community G_1 and get the community G_1. Repeat the above process, and get the other 10 communities.

There are 11 communities according to our algorithm, but lack the team composed of node 36, 42, 80, 82, 90. In fact, this team represents Independents. The matches between Independents are little, the edges between these five nodes are less, so the Independents should not be counted as a single community. The community structure by this algorithm is shown in Figure 5.

Figure 5. College Football Network Obtained by this Algorithm

By experiment, the results are the roughly the same when the belonging degree b_1 is in (0.25,0.4) and b_2 is in (0.4,0.5). The correct rate of the GN [4], LP [14], FCM [8] and this algorithm is shown in Table 1.

Table 1. Contrast 4 Algorithms Accuracy

algorithm	Number of communities	accuracy /(%)
GN	11	78
LP	11	87
FCM	10	90
Algorithm in this paper	11	88

4. Conclusion

In this paper, we present a new method to divide community structure quickly in complex network based on node belonging degree with respect to the community. Compared with other methods, this method does not need to know the complexity of the network, the number of nodes and the number of communities in network. When we want to know which community the node belongs to (such as Zachary Karate club network), then the algorithm is simpler. Meanwhile, this method reduces the time and space complexity because it only needs to know the local message of the network. For a network with three communities with 19 nodes, Zachary Karate club network and College Football network, the experiment results show that this method is feasible.

References

[1] Wang Xiao-fan, Li Xiang. Chen Guan-rong. Theory of complex networks and its application (in Chinese). Beijing: Tsinghua University Press. 2006: 162-191.
[2] Kemighan BW, Lin S. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal.* 1970; 49(2): 291-307.
[3] Wu F, Huberman BA. Finding communities in linear time: a physics approach. *Eur Phys J B*. 2004; 38: 331-338.
[4] Girvan M, Newman MEJ. Community Structure in Social and Biological Networks. *P Natl Acad Sci USA*. 2002; 99(12) :7821-7826.
[5] Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E*, 2004; 69(2) : 026113.
[6] Palla G, Dernyi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society. *Nature.* 2005; 435(7043): 814-818.
[7] Kumpula JM, Kivel M, Kaski K, et al. Sequential Algorithm for Fast Clique Percolation. *Physical Review E*. 2008; 78(2): 026109.
[8] Zhang S, Wang R, Zhang X. Identification of Overlapping Community Structure in Complex Networks Using Fuzzy C-means Clustering. *Physica A: Statistical Mechanics and it is Applications*. 2007; 374(1): 483-490.
[9] Lancichinetti A, Fortunato S, Kert esz J. Detecting the Overlapping and Hierarchical Community Structure in Complex Networks. *New Journal of Physics*. 2009; 11: 033015.
[10] Newman ME, Leicht EA. *Mixture Models and Exploratory Analysis in Networks*. Proc Natl Acad Sci USA. 2007; 104(23): 9564-9569.
[11] EvansT, Lambiotte R. Line Graphs, Link Partitions, and Overlapping Communities. *Physical Review E.* 2009; 80(1): 16105.
[12] Ahn YY, Bagrow JP, Lehmann S. Link Communities Reveal Multiscale Complexity in Networks. *Nature*. 2010; 466: 761-764.
[13] A Capocci, VDP Servedioa BG, Caldarellia bF Colaiori. Detecting communities in large networks. *Physica A.* 2005; 352: 669–676.
[14] Agarwal G, Kempe D. Modularity-maximizing graph communities via mathematical programming. *Eur Phys J B*. 2008; 66: 409-418.
[15] S Chanda, A De. Congestion relief of contingent power network with evolutionary optimization algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(1): 1-8.
[16] BW Yohanes, Handoko, HK Wardana. Focused crawler optimization using genetic algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2011; 9(3): 403-410.