

# A Local Community Detection Algorithm for Large Scale Wireless Networks

Dong Jia

School of Telecommunications Engineering, Xidian University, Xi'an, China  
Email: dongjia09@foxmail.com

**Abstract**-Maintaining global dynamic routing in large-scale wireless network requires lots of overhead due to the mobility of nodes, thus the hierarchical network structure is widely used in reality. Therefore, this paper proposes a modified local community detection algorithm in large scale wireless network scenario to enhance the network self-organization. In this paper, we first describe the scenario of large-scale wireless network communication. Then the algorithm of local community detection is adopted to cluster the nodes that communicate with others more frequently into a community. Finally the local community detection algorithm is modified with a selection factor. The simulation results show that the new algorithm has higher detection accuracy than classic local community detection algorithms.

**Keywords**-wireless network; local community detection; hierarchical network

## I. INTRODUCTION

Wireless ad hoc networks and other large-scale wireless networks are widely used in internet of vehicles, military and commercial systems with the characteristics of limited bandwidth, easy mobility of nodes and changing topology [1]. Due to the mobility of nodes, maintaining global dynamic routing requires a lot of control information[2]. The larger the network size, the greater the cost of route maintenance. Local community detection algorithm is used to detect the connected nodes started from the source node, then wireless networks can be divided into communities or clusters [3] using the local community detection algorithm.

The community is a concept in social network that is considered to be a collection of nodes whose inter-connections within a community are much larger than the number of connections to the outside of the community. Community detection is essentially a method of topological clustering, through which connected nodes in the network can be divided into communities.

For community detection, there have been some classic algorithms, such as the earlier proposed GN[4] algorithm and modified algorithm BGLL[5]. These algorithms have been able to partition the community structure well, and the modified algorithm is also more efficient. The defect is that they are global community detection algorithms. In large scale wireless network scenario, If the community structure is expected to be available starting from a specific node, the local community detection algorithm that progressively scale out need to be used.

Local community detection is a local optimization algorithm, so there will be local optimization problems in practical application. In order to mitigation the local

optimization problem, a modified local community detection algorithm is proposed.

In this paper, we discuss the large-scale wireless network scenarios in which a source node detects its surrounding nodes by point-by-point diffusion. According to the traffic between each other, nodes that communicate frequently with each other are clustered into the same community.

## II. NETWORK MODEL

As shown in Figure 2.1, we consider a large-scale wireless network scenario where nodes communicate with each other at a time. it is an abstracted and unweighted network. Each node is a terminal in the wireless network. The edge connection indicates that the communication frequency and the traffic exceed a certain threshold. If the local community detection algorithm is used to find a community from a source node, the wireless network will becomes a hierarchical structure. This method is different from the previous method with the gateway node [6], especially it is more self-organized.

In this paper, we consider that a node has strong connections with others when the communication frequency and the traffic being greater than a certain threshold. In this way, a large-scale wireless network is abstracted into a network with nodes and edges. The advantage of the model is that nodes that communicate with each other more frequently can be clustered as a local community. So that the routing overhead of the network is reduced.

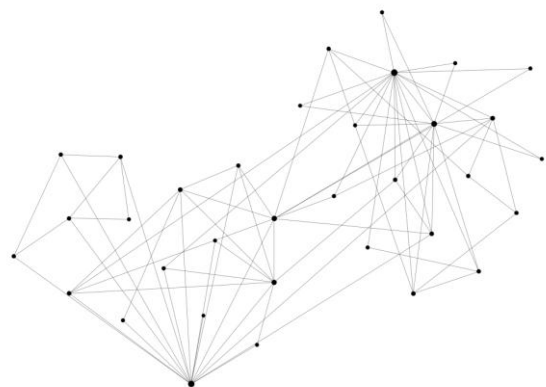


Figure 1. Network model

## III. EXISTING COMMUNITY DETECTION ALGORITHM

The classic local community detection methods have been proposed by Clauset [7] and by Luo [8]. Both of these

methods are based on maximizing the local modularity, with the definition of the local modularity and the way the algorithm handled are different. The method proposed by Luo divides the degree of internal nodes by the degree of external nodes to define local modularity. The algorithm is divided into two stages: addition and deletion. Clauset method is concerned with the boundary which has connections with inside and outside the community. The local modularity is defined with the boundary, and a node is added to the community that maximizes the local modularity each time.

#### A. Degree - Based Local Community Detection Algorithm

The in-degree is extended defined as the sum of the number of connections of nodes within the same community, and the out-degree is the sum of the number of connections with other communities. Based on the new degree definition, the network has a weak modularity means that the sum of in-degree in the sub-graph more than the sum of out-degree in the sub-graph.

Given  $G$ , the adjacency matrix is defined as:

$$A_{ij} = \begin{cases} 1 & \text{nodes } i \text{ and } j \text{ have connections,} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

The sub-graph  $S$  is known,  $S$  and its adjacent node set:

$$S_{ij} = \begin{cases} 1 & \text{nodes } i \text{ and } j \text{ are connected} \\ & \text{and at least one is in } S, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Based on the above definition, the in-degree of sub-graph can be defined as:

$$ind(S) = \sum_{i,j} S_{ij} \delta(i, j) \quad (3)$$

Where the value of  $\delta(i, j)$  is 1 if both  $i$  and  $j$  are inside the sub-graph  $S$ , 0 otherwise. The out-degree is defined as:

$$outd(S) = \sum_{i,j} S_{ij} \lambda(i, j) \quad (4)$$

Where the value of  $\lambda(i, j)$  is 1 if and only if one of them is inside sub-graph  $S$ , and 0 otherwise.

Using the weak modularity evaluation [9], the definition of local modularity is:

$$M = \frac{ind(S)}{outd(S)} \quad (5)$$

The algorithm is divided into two stages, the adding stage and the deleting stage. Place the source node first, and then place the source node's adjacent point. The adding stage is

put into nodes which increase the local module degree most in turn; after the adding stage, the deleting stage deletes nodes which decrease the local module degree most.

#### B. Boundary - Based Local Community Detection Algorithm

Suppose that in the graph  $G$ , we are very clear about the connectivity of some of these nodes, these known parts, we call  $C$ . This means that there is another part of the region  $U$ , we only know that nodes in this region have adjacent nodes in  $C$ . More information can only be probed start from the source node. If a node is added as a member according to certain rules, then its neighbor is added to  $U$ . As shown in Figure 2:

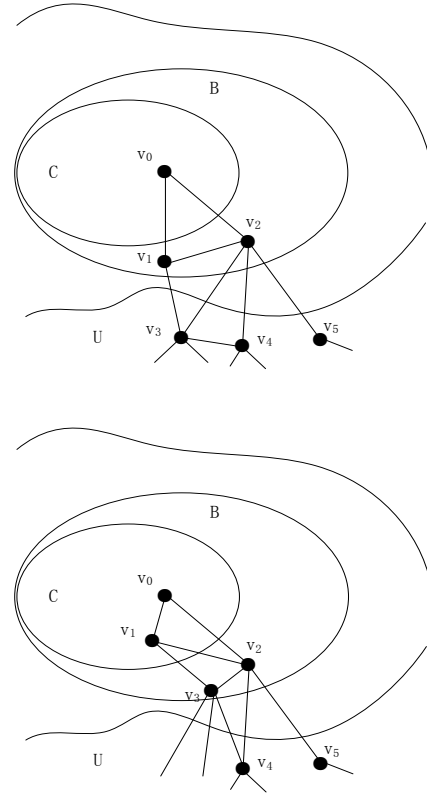


Figure 2. Local community detection

The adjacency matrix is defined as:

$$A_{ij} = \begin{cases} 1 & \text{nodes } i \text{ and } j \text{ are connected} \\ & \text{and at least one is in } C, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Considering the known part of the local community network, the simplest criterion for evaluating the quality of network partitioning is the proportion of the number of edges falling completely inside, as follows:

$$\frac{\sum_{ij} A_{ij} \xi(i, j)}{\sum_{ij} A_{ij}} = \frac{1}{2m} \sum_{ij} A_{ij} \xi(i, j) \quad (7)$$

Where  $m$  is the total number of edges of the entire network, and  $\xi(i, j)$  is 1 if the nodes  $i$  and  $j$  are both in  $C$ , and 0 otherwise.

If nodes in  $C$  have at least one connection with nodes in  $U$ , we define this set as the boundary  $B$ .

Where the boundary set  $B$  is determined by:

$$B_{ij} = \begin{cases} 1 & \text{nodes } i \text{ and } j \text{ are connected} \\ & \text{and at least one is in } B, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

For each adjacent node in the set of  $B$ , calculate the local modularity added by adding the neighbor to the community:

$$R = \frac{\sum_{ij} B_{ij} \delta(i, j)}{\sum_{ij} B_{ij}} = \frac{I}{T} \quad (9)$$

$\delta(i, j)$  is defined as:

$$\delta(i, j) = \begin{cases} 1 & \text{both nodes } i \text{ and } j \text{ are in } C, \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Where  $I$  denotes the number of connections having an endpoint in  $B$  and no endpoint in  $U$ , and  $T$  is the sum of the number of connections which have at least one endpoint in the set of boundaries.

According to the above definition, the value of  $R$  is clearly between 0 and 1, because of  $T \geq I$ . While the degree based local community detection algorithm has that value greater than 1, then the modification of this paper is based on classical boundary based local community detection algorithm. We refer to the modified local community detection algorithm as the Mlocal.

#### IV. MODIFIED LOCAL COMMUNITY DETECTION ALGORITHM

Because the boundary based local community algorithm has local optimization problem, it is not to select a new node which has the maximum addition of local modularity directly, but to select using a random factor[10].

The added value of modularity:

$$\Delta R_j = \frac{a - bR - c(1 - R)}{1 + T - c + b} \quad (11)$$

Where  $a$  is the sum of the number of edges connected to  $v_j$  and the nodes in the community.  $b$  is the added value of

$T$  due to the addition of node  $v_j$ .  $c$  is the number of edges that need to be removed from  $T$  due to the addition of nodes.

A selected metric is determined for the node based on the modularity increment for each node. The metric for the node in the neighbor set is:

$$S_j = \frac{\Delta R_j}{\sum_i \Delta R_i} \quad (12)$$

Each node randomly generates a selection factor, and each node added to the community is determined according to the selection factor, instead of directly selecting the nodes that maximize the local modularity. The selection factor is a random number between 0 and 1, and the selected node number is the minimum value  $j$  that satisfies the following equation:

$$\sum S_j > rand \quad (13)$$

The meaning of the left type is the sum of the selected metric from the first to the  $j$  node.

#### V. PERFORMANCE COMPARISON OF LOCAL AND MLOCAL

The simulation utilizes the LFR benchmark network, which simulates the node degree and community size in real networks, and is the most commonly used simulation data set in current community detection research. The LFR benchmark network has several parameters that can be set to different parameters to generate different types of analogue networks. The LFR network has a known community structure that can be used to evaluate the quality of the communities found by the algorithm. Different types of analog networks can be generated using different parameters.

Because it is a local community detection algorithm, precision and recall, which are widely used in information retrieval and statistical classification, are used to evaluate the quality of the results, as well as the combined metric  $F$ .

LFR network parameter settings:

TABLE I. LFR PARAMETERS

Parameter	Meaning	Value
N	number of nodes	400
k	average degree	12
maxk	maximum degree	25
mu	mixing parameter	0.1
minc	minimum for the community sizes	20
maxc	maximum for the community sizes	35

The simulation condition is to select one source node at a time, and then run the local community detection algorithm and the modified algorithm separately, adding 30 nodes at a time, which is larger than the average node number of the community by 27.5. The performance of the algorithm is related to the selection of the source node, then several

experiments are performed to select different source nodes. The final result is the average performance of algorithms.

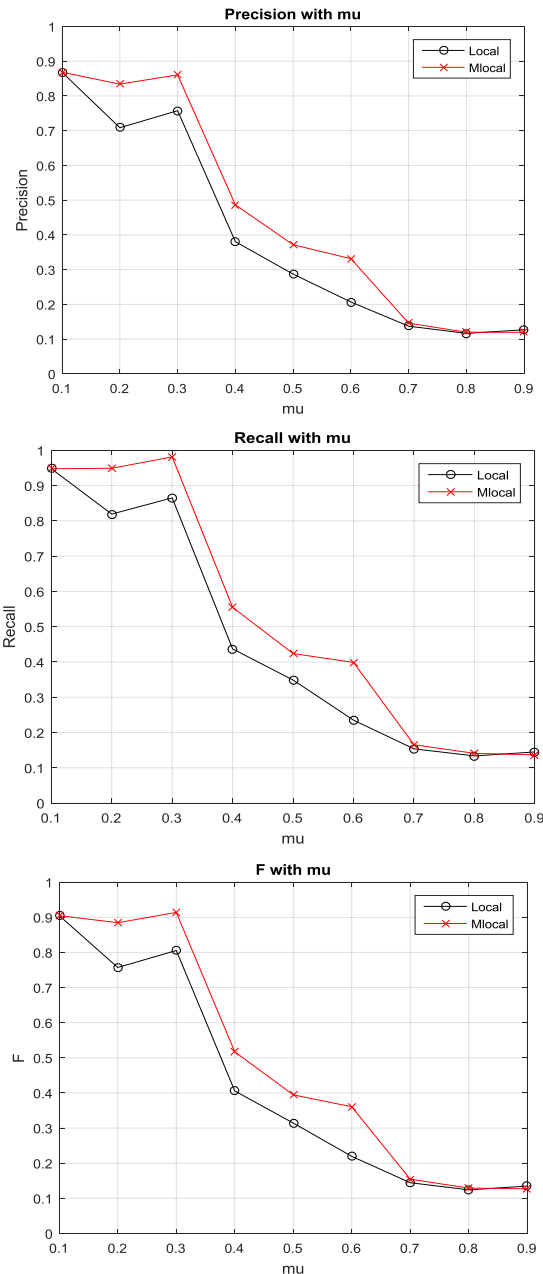


Figure 3. Accuracy comparison

As shown in Fig.3, the detection accuracy can be improved using the modified local community detection algorithm.

As mentioned above, the performance of the algorithm has relevance to the selection of source, then the correct detection probability of the two algorithms can be verified by simulation. The specific method is to run community detections repeatedly with each node as a source node. The probability of correct detection can be obtained by dividing the number of correct detections by the total number of nodes.

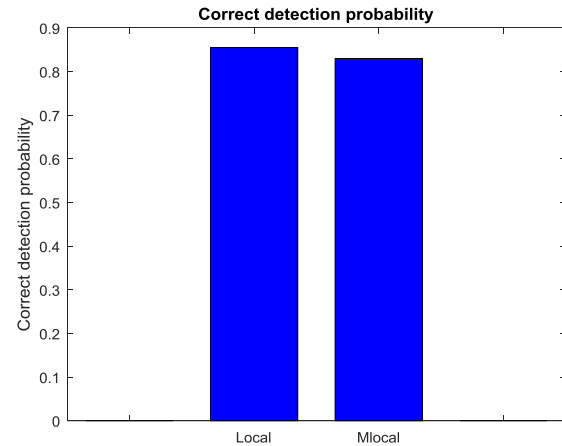


Figure 4. Correct detection probability

As shown in Fig.4, while Mlocal found an increase in detection accuracy, the probability of correct detection decreased. Therefore, the algorithm based on the actual scene needs to be selected according to network conditions.

## VI. SUMMARY

By applying the modified local community algorithm to large-scale wireless networks, the local clustering structure of the network can be found and wireless network can be more self-organized and flexible than the previous algorithms. The modified algorithm is a local community detection algorithm based on the selection metric, which introduces a random factor, then the local optimal problem can be alleviated to a certain extent. How to determine the threshold of communication between nodes and how to use local community detection algorithm to reduce the route cost is the future research direction.

## REFERENCES

- [1] P. Samar and S. Wicker, "Link Dynamics and Protocol Design in a multihop Mobile Environment," *IEEE Transactions on Mobile Computing*, vol. 5, no. 9, pp. 1156-1172, Sep.2006.
- [2] D. Wang and A. Abouzeid, "Link State Routing Overhead in Mobile Ad Hoc Networks: A Rate-Distortion Formulation," *IEEE INFOCOM 2008*, pp. 2011-2019, May. 2008.
- [3] L. Qing, Q. Zhu and M. Wang, "Design of a distributed energy efficient clustering algorithm for heterogeneous wireless sensor networks," *IEEE Wireless Communications*, Vol. 29, no. 12 .pp. 2230-2237, Aug 2006.
- [4] M. Newman and M. Girvan, "Finding and Evaluating Community Structure in Networks," *Physical Review E Statistical Nonlinear and Soft Matter Physics*, E. 69, 2004.
- [5] V. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre, "Fast Unfolding of Communities in Large Networks," *Journal of Statistical Mechanics: Theory and Experiment* , pp.155-168, Oct 2008.
- [6] X. Wang, L. Qian and J. Wu, "An Energy and Distance Based Clustering Protocol for Wireless Sensor Networks," *Novel Algorithms and Techniques in Telecommunications and Networking*, pp. 409-412, 2010.
- [7] A. Clauset, "Finding local community structure in networks," *Physical Review E Statistical Nonlinear and Soft Matter Physics*, Vol. 72, no. 2 .pp. 254-271, 2005.

- [8] F. Luo, J. Wang and E. Promislow, "Exploring Local Community Structures in Large Networks," Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Vol. 6, no. 4, pp. 387-400, 2008.
- [9] R. Shang, J. Bai and L. Jiao, "Community detection based on modularity and an modified genetic algorithm," Physica A, vol. 392, no. 5, pp. 1215-1231, 2013.
- [10] X. Zhang, L. Wang and Y. Li , "Extracting local community structure from local cores," Proceedings of the 16th International Conference Database Systems for Advanced Applications, Vol. 6637, pp. 287-298, 2011.