

An Overlapping Communities Detection Algorithm via Maxing Modularity in Opportunistic Networks

Zhi-Peng GAO¹,Jing-Jing ZHANG^{1,a} and Kun NIU²

¹State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China

²School of Software Engineering, Beijing University of Posts and Telecommunications, Beijing, China

^a Corresponding author: jingjing_zhang1992@163.com

Abstract. Community detection in opportunistic networks has been a significant and hot issue, which is used to understand characteristics of networks through analyzing structure of it. Community is used to represent a group of nodes in a network where nodes inside the community have more internal connections than external connections. However, most of the existing community detection algorithms focus on binary networks or disjoint community detection. In this paper, we propose a novel algorithm via maxing modularity of communities (MMC)to find overlapping community structure in opportunistic networks. It utilizes contact history of nodes to calculate the relation intensity between nodes. It finds nodes with high relation intensity as the initial community and extend the community with nodes of higher belong degree. The algorithm achieves a rapid and efficient overlapping community detection method by maxing the modularity of community continuously. The experiments prove that MMC is effective for uncovering overlapping communities and it achieves better performance than COPRA and Conductance.

1 Introduction

Opportunistic networks[1] are special networks in which nodes contact each other opportunistically to forward information. Due to unpredictable node mobility and without any fixed infrastructures, there is not an end-to-end path in most situations. Different from the store-and-forward manner in traditional networks, information are forwarded in a store-carry-and-forward manner, so applications need to tolerant long period of time delay in opportunistic networks. For example, people can use portable intelligent devices with short-range wireless communication capability (e.g. Bluetooth, WiFi) or some computing power to store and forward information, it can forward information more conveniently and easily without network infrastructure.

Community detection in opportunistic networks has become a significant and hot issue, which is used to understand characteristics of network through analyzing structure of it. Community is used to represent a group of nodes in opportunistic networks where internal connections of nodes inside the community are denser than external connections. Community detection can help us to uncover and understand local community structure in both offline mobile trace analysis and online applications, and it is helpful in decreasing forwarding time as well as the storage capacity of nodes. Since the relationships between nodes usually seem to be stable and less volatile than node mobility, forwarding schemes

based on community [2-6]outperform traditional approaches [7, 8]. Overlapping community detection, one of the most interesting research of community detection, is the primary focus of this paper. Overlapping community means that a node may participate in more than one community in the network. Furthermore, most of real-world networks exhibit the feature of overlapping communities, such as social networks, information networks and biological networks. For example, we divide communities according to people's interests. A man may belong to multiple communities. People who like sports may have interests in music and others may still have interests in cooking.

The rest of this paper is organized as follows. Section 2 introduces related work in community detection. Section 3 presents our community detection algorithm for opportunistic networks. Then, we evaluates the performance of our proposed algorithms with COPRA and Conductancein Section 4. Section 5 concludes the paper and states the future of the field.

2 Related Work

Several community detection algorithms have been proposed for opportunistic networks. In this section, we divide the community detection algorithms into three categories: modularity-based, label propagation-based and attribute-based.

In modularity-based detection algorithms, early scheme is GN algorithm [9]. Girvan and Newman proposed edge betweenness which means the number of shortest paths in which the given edge is included. In this method, the edges with high edge betweenness score will be removed in every step. However, the method needs to recalculate the score of edge betweenness for all edges every time after each removals. It is computationally intensive and suffers with scalability problem. Additionally, Newman et al. [10] proposed a bottom-up hierarchical approach that optimizes the modularity score in a greedy manner. Initially, every node is a community. Communities are merged iteratively based on optimal modularity score until there is no increment in modularity score. In [4], Pan Hui et al. proposed a detection algorithm based on cliques and modularity. In addition, CPM is proposed in [11], which is based on the assumption that a community consists of all k-cliques that can be reached from each other through a series of adjacent k-cliques. A k-clique is a fully connected subgraph and two k-cliques are said to be adjacent if they share k-1 nodes. However, it doesn't consider all the characteristics of links, such as connection time or connection frequency. The value of k is also hard to be determined and the method is more suitable for networks with dense connected parts.

In label propagation-based detection algorithms, the typical algorithm is the Label Propagation Algorithm (LPA) [12], researchers present an algorithm that assigns k labels to each node in the network and updates its label according to the most frequent label in its neighborhood. This method is faster than others but produces different results each time based on initial configuration. So one need to run the algorithm several times to build the consensus. It consumes time. SLPA[13] is an extension of LPA, in this method each node has a memory and considers information that has been observed in the past to make current decision. COPRA[14] can achieve good performance in some cases, but it limits the number of communities for each node which decreases the accuracy whenever too big or too small. In [15], authors propose a balanced multi-label propagation algorithm (BMLPA) for overlapping community detection. Compared to COPRA, the advantage of the strategy is that it allows nodes to belong to any number of communities without a global limit.

In [16], researchers present an Interest Community Routing (ICR) algorithm which is founded on social network theory. Authors define an interest metric and a message header to represent individual interests and data types in the networks. By comparing the similarity between the message header and the interest metric of node, the node will be put into corresponding interest community. However, the interest is not stable and one may have many kinds of interests in reality, so it's not applicable.

These methods only detect disjoint communities, or are applied in networks where nodes are in frequent contacts, or need to run many times to achieve stable status. Those are not sufficient to process opportunistic networks with overlapping communities effectively.

In this paper, we first calculate relation intensity between nodes in opportunistic networks, it can solve the problem which community structure is fuzzy in binary networks. In binary networks, multiple contacts and single contact are considered the same as having contacts. We consider nodes with higher relation intensity as initial community. Finally, we extend the community with nodes of higher degree so that the modularity of community can be increased continuously while solve the problem of overlapped community detection.

3 MMC Algorithm

We aggregate node mobility traces into weighted contact graphs. The vertices of the graphs are nodes, the edges are relationships, and the weight of edge is computed based on the times of contacts and the duration of each contact between two nodes. Like social networks, the more often they contact, the more familiar they are. And the more time they spend together, the closer they are. For simplicity, we use $G = (V, E)$ to represent weighted contact graphs, where V is the set of nodes and $E\{(v, w)\}$ is the set of edges. We use the adjacency matrix A to represent the edge in E and A_{vw} to denote the weight of the edge between two nodes $v, w \in V$.

3.1 Definitions

3.1.1 Relation Intensity

In order to help us present the contact information between two nodes and make it easier for further processing, we simplify the presentation of contact record in [19]. We use a three tuples $g_v^s = \{id, t_s, t_e\}$ to simply represent a contact between two nodes in opportunistic networks. g_v^s is the s th contact record of node v , id is the node which v have contact with, t_s represents the time that contact starts and t_e represents the time that contact ends. Thus, node v can get n contact records after a short time which we can show like $G_v = \{g_v^1, g_v^2, g_v^3, \dots, g_v^n\}$, where G_v is the contact history of node v . So we can get the following equations to calculate the relation intensity between node v and any other node w .

$$fre_v^w = \frac{|G_v^w|}{|G_v|} \quad (1)$$

$$int_v^w = \frac{\sum_{g_v^s \in G_v^w} (t_e - t_s)}{\sum_{g_v^s \in G_v} (t_e - t_s)} \quad (2)$$

Where $G_v^w = \{g | g_v^s \in G_v, id = w\}$. fre_v^w is the contact frequency, int_v^w is the fraction of time that node v spend with node w over the time that node v spend with all nodes.

The relation intensity between node v and node w is shown in Eqs(3):

$$A_{vw} = \left[(\alpha fre_v^w)^2 + (\beta int_v^w)^2 \right]^{\frac{1}{2}} \quad (3)$$

$$s.t. \alpha \in [0,1], \beta \in [0,1], \alpha + \beta \leq 1$$

Where α and β are the weighting factors. Values of α and β as the case may be can be changed considering the proportion between the number of contacts and the duration of each contact.

3.1.2 Modularity

In this paper, we use Newman's weighted modularity proposed in [4] as a measurement of the quality of the community structure it detects. One can get the corresponding fitness value with the following definition of modularity (Q):

$$Q = \sum_{vw} \left[\frac{A_{vw}}{2m} - \frac{k_v k_w}{(2m)^2} \right] \delta(c_v, c_w) \quad (4)$$

$$m = \frac{1}{2} \sum_{vw} A_{vw} \quad (5)$$

$$\delta(c_v, c_w) = \begin{cases} 1, & c_v = c_w \\ 0, & c_v \neq c_w \end{cases} \quad (6)$$

where A_{vw} is the relation intensity between node v and w , if such an edge exist, and 0 otherwise; k_v is the degree of v defined as the total weight of edges connected with v , thus $k_v = \sum_v A_{vw}$; m is the sum weight of all edges in the network; and c_v denotes the community which node v belong to. The network community structure is represented by $C = \{C_1, C_2, C_3, \dots\}$, where $C_v \in C$. We do not require $C_v \cap C_w = \emptyset$, as we detect overlapping community here.

From the formula (4), Q is defined as this fraction minus the fraction of the edges that would be expected to fall within the communities if the edges were assigned randomly while we keep the degrees of the vertices unchanged. Generally, if Q is greater, community structure is clearer; on the contrary, if Q is smaller, community structure is more ambiguous.

3.1.3 Belonging Degree

During the detection of community structure, the belonging degree proposed in [17] between a node v and a community C is defined as:

$$B(v, C) = \frac{\sum_{w \in C} A_{vw}}{k_v} \quad (7)$$

If all neighbors of a node v are included in community C , $B(v, C) = 1$.

3.2 The Community Detection Algorithm

Our detection algorithm works as follows:

Step 1: After a period of warm time t , each node begin to calculate the relation intensity between itself and other nodes according to the contact history. We get a weighted contact graph.

Step 2: Sort edges of contact graph by size. Then choose two nodes with highest relation intensity as a new community C and calculate its modularity according to Eqs.(4).

Step 3: Expand C , put neighbors of nodes in C into N_C and sort them by belonging degree. Choose the node with the highest degree to C to form a new community C' , calculate Q of C' . Then there are two situations need to be handled:

If Q increases, repeat from Step 3 and the expanding process is continued for community C' ;

Otherwise, the expanding process of C is finished. Meanwhile, the edges within community C are removed from the edge set E .

Step 4: Repeat from Step 2 to Step 3 until E is empty. For completeness, the pseudo code of the detecting algorithm is shown as follows.

```

Input: traces of nodes and nodes set V
Output: Community of node
1:  $G = (V, E) \leftarrow$  compute relation intensity
   between nodes of V;
2: sort edge  $e \in E$ ;
3: foreach edge  $e \in E$ 
4:    $e.visited = false$ ;
5: end foreach
6: foreach edge  $e \in E$ 
7:   if  $e.visited = false$ ;
8:      $C \leftarrow v, w$ ;
9:      $N_C \leftarrow Neighbours(v) \cup Neighbours(w)$ ;
10:    foreach node  $u \in N_C$ 
11:      compute belongDegree(u);
12:    end foreach
13:    sort  $u \in N_C$ ;
14:    foreach node  $u \in N_C$ 
15:      if  $\nabla Q > 0$ 
16:         $C \leftarrow u$ ;

```

```

17:       $N_C \leftarrow Neighbours(u) \cup N_C$ ;
18:      goto (10);
19:      else
20:        complete detection of  $C$ ;
21:      end for each
22:      foreach edge  $e$  between nodes in  $C$ 
23:         $e$ . visited = true;
24:      end if
26: end foreach
27: return Community of node;

```

According to formula (4), if Q increase, node u is involved in community C , we have:

$$\nabla Q = \frac{1}{2m} \left(\sum_{v \in (C \cap N_u)} A_{uv} - \frac{k_u \sum_{v \in (C \cap N_u)} k_v}{2m} \right) > 0 \quad (8)$$

From formula (7), we have:

$$B(v, C) > \frac{\sum_{v \in (C \cap N_u)} k_v}{2m} \quad (9)$$

Node u will be added into Community C when (9) is founded.

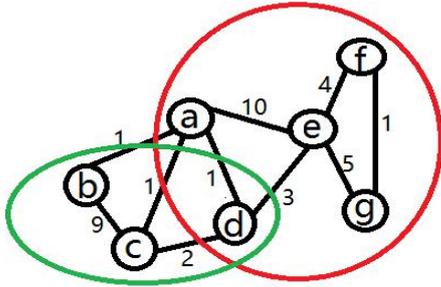


Figure 1.An example for community detection.

We use an example to illustrate the expanding process. Fig.1 shows an example of overlapping community detection using our detection algorithm. Two communities are detected and shown in different circles. As shown in Fig.1, we consider $C = \{a, e\}$ as initial community, neighbors of nodes in C is $N_C = \{b, c, d, f, g\}$. The belonging degree of nodes in N_C is $\{1/10, 1/12, 2/3, 4/5, 5/6\}$. We put node g with the highest value into C and we get $C' = \{a, e, g\}$. Because $\nabla Q > 0$, the expanding process is continued for newly formed C , and now $N_C = \{b, c, d, f\}$. The belong degree is $\{1/10, 1/12, 2/3, 1\}$. Thus, node f is added into C directly. Repeat the expanding process until we get $C' = \{a, e, g, f, d\}$. Now $N_C = \{b, c\}$, and the

belonging degree is $\{1/10, 1/4\}$. We try to put node c into the community and compute ∇Q . Since $\nabla Q < 0$, the expanding process of current community is finished. We remove edges in this community from E . Then we choose $C = \{b, c\}$ as the new community. The expanding process is continued for new community. Finally, we get another community $C = \{b, c, d\}$. Remove the edges in new community, and the algorithm is ended because E is empty. We get two communities: $\{a, e, g, f, d\}$ and $\{b, c, d\}$. The overlapping node is $\{d\}$.

4 Experiments

In this section, we present some representative numerical results to validate the effectiveness of MMC comparing with two other schemes: Conductance [17] and COPRA [14], both of which can detect overlapping communities.

We use LFR benchmark [18] proposed by Lancichinetti et al. to test overlapping community detection algorithms. LFR is a well-known benchmark which can generate synthetic unweighted or weighted networks. It provides power-law distributions of node degree and community size, allows overlaps between communities. We can set many parameters to control the generated network, such as the fraction of nodes that belong to more than one community over all nodes and the number of communities that a node belongs to simultaneously. And the settings of parameters we set are shown in Table 1.

Table 1. Parameter Settings in LFR.

Parameter	Value	Parameter	Value
N	1000, 4000	γ	0~0.5
mut	0.1, 0.3	muw	0.1, 0.3
ϵ	1.5	k	20, 40
k_{max}	50, 100	τ_1	2
τ_2	2	o_m	2

The meanings of all these parameters are as follows:

N is the number of all nodes while γ is the fraction of overlapping nodes over all nodes. mut and muw denote mixing parameter for topology and mixing parameter for edge weights respectively. The exponent for weight distribution is ϵ . k and k_{max} represent the average of node degree and the maximum of node degree respectively.

τ_1 is minus exponent for degree sequence while τ_2 is minus exponent for community size distribution. In addition, the number of communities that overlapping nodes belong to is denoted by O_m .

We use above metrics and different parameter settings: $mut = muw = 0.1, mut = muw = 0.3, N=1000,4000$. And the overlapping fraction γ is changed from 0 to 0.5 for each setting to evaluate the performance of these algorithms. Fig. 2 to Fig.5 show the experimental results of three algorithms: COPRA, Conductance, and MMC.

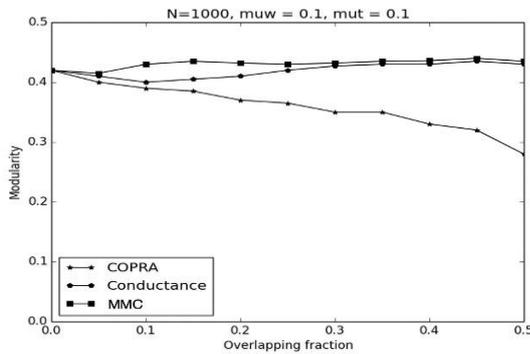


Figure 2. Modularity of the community structure detected by COPRA, Conductance and MMC while $N=1000$ and $muw=mut=0.1$.

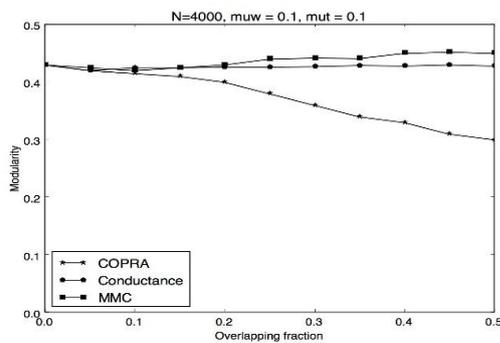


Figure 3. Modularity of the community structure detected by COPRA, Conductance and MMC while $N=4000$ and $muw=mut=0.1$.

Fig.2 and Fig.3 show modularity of the community structure detected by three algorithms while $mut=muw=0.1$ and number of nodes is different. As is shown, the modularity of all algorithms is higher. This may be because network structure is not very complex. For both COPRA and Conductance, the modularity is lower than MMC. In addition, the modularity of COPRA dramatically decreases while the overlapping fraction increases in all settings.

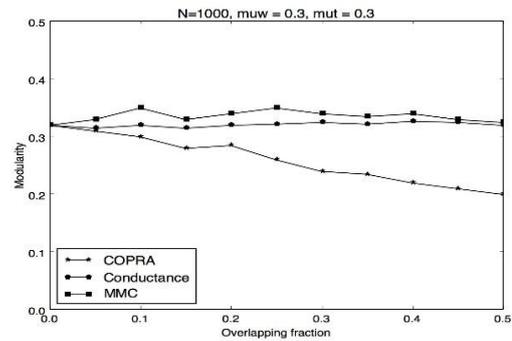


Figure 4. Modularity of the community structure detected by COPRA, Conductance and MMC while $N=1000$ and $muw=mut=0.3$.

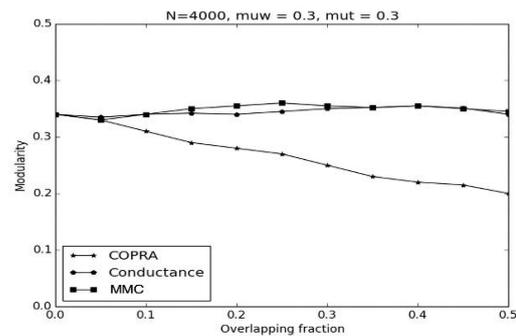


Figure 5. Modularity of the community structure detected by COPRA, Conductance and MMC while $N=4000$ and $muw=mut=0.3$.

Fig.4 and Fig.5 show modularity of the community structure detected by three algorithms for while $mut=muw=0.3$ and number of nodes is different. When mut and muw is higher, and the number of nodes do not change, the modularity of all algorithms is lower. The performance of COPRA is still worse while γ increases. In real situation, it's hard to determine a global limit γ which has a great influence on performance for COPRA. However, MMC is stable with the increase of the overlapping fraction γ . Thus, we can conclude that MMC has the ability to detect overlapping communities.

5 Conclusions

In this paper, we propose an overlapping community detection algorithm for opportunistic networks. First, we introduce relation intensity to measure relationships among nodes. Then we design a new detection algorithm based on maxing modularity of communities. Simulation results on synthetic networks show that our community detection algorithm is much better than other algorithms in terms of modularity. In future, we will do more work on the validations in different environments. We will evaluate this algorithm on large scale networks coming from real networks.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant no. 61272515, 61372108, 61401033), and National Science & Technology Pillar Program (Grant no. 2015BAH03F02).

References

1. Huang, C. M., Lan, K. C., & Tsai, C. Z. *A survey of opportunistic networks*. In Advanced Information Networking and Applications-Workshops. 22nd International Conference on IEEE (pp.1672-1677)(2008).
2. Daly, E. M., & Haahr, M. *Social network analysis for routing in disconnected delay-tolerant manets*. In Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing (pp. 32-40) (2007).
3. Hossmann, T., Spyropoulos, T., & Legendre, F. *Know thy neighbor: Towards optimal mapping of contacts to social graphs for dtn routing*. In INFOCOM, 2010 Proceedings IEEE (pp. 1-9)(2010).
4. Hui, P., Crowcroft, J., & Yoneki, E. *Bubble rap: Social-based forwarding in delay-tolerant networks*. IEEE Transactions on Mobile Computing, **10**(11), 1576-1589(2011).
5. Gao, W., Li, Q., Zhao, B., & Cao, G. *Social-aware multicast in disruption-tolerant networks*. IEEE/ACM Transactions on Networking (TON), **20**(5), 1553-1566(2012).
6. Gao, W., et al., *On Exploiting Transient Social Contact Patterns for Data Forwarding in Delay-Tolerant Networks*. IEEE Transactions on Mobile Computing, **12**(1): p. 151-165(2013).
7. Becker, V. D. *Epidemic routing for partially connected ad hoc networks*. Proceedings of Technique Report, Department of Computer Science, Duke University, Durham, UK (2000).
8. Yuan, Q., Cardei, I., & Wu, J. *Predict and relay: an efficient routing in disruption-tolerant networks*. In Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing (pp. 95-104)(2009).
9. Girvan, M. and M. Newman, *Community structure in social and biological networks*. Proceedings of the national academy of sciences, **99**(12): p. 7821-7826(2002).
10. Newman, M.E. and M. Girvan, *Finding and evaluating community structure in networks*. Phys Rev E Stat Nonlin Soft Matter Phys, **69**(2 Pt 2): p. 026113(2004).
11. Palla, G., et al., *Uncovering the overlapping community structure of complex networks in nature and society*. NATURE, 435(7043): p. 814-818(2005).
12. Raghavan, U.N., R. Albert and S. Kumara, *Near linear time algorithm to detect community structures in large-scale networks*. Phys Rev E Stat Nonlin Soft Matter Phys, **76**(3 Pt 2): p. 036106(2007).
13. Xie J, S.B.K.L., *SLPA: Uncovering Overlapping Communities in Social Networks via a Speaker-Listener Interaction Dynamic Process*. (2011).
14. Gregory, S., *Finding overlapping communities in networks by label propagation*. NEW JOURNAL OF PHYSICS, **12**(103018)(2010).
15. Wu, Z., et al., *Balanced Multi-Label Propagation for Overlapping Community Detection in Social Networks*. Journal of Computer Science and Technology, **27**(3): p. 468-479(2012).
16. Liu, Q., et al., *An Interest Community Routing Scheme for Opportunistic Networks*, in IEEE Global Telecommunications Conference (Globecom). p. 4366-4371(2013).
17. Lu Z, Sun X, Wen Y, et al. *Algorithms and applications for community detection in weighted networks[J]*. Parallel and Distributed Systems, IEEE Transactions on, **26**(11): 2916-2926(2015).
18. Lancichinetti A, Fortunato S. *Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities[J]*. Physical Review E, **80**(1): 016118(2009).
19. Ma, X. and X. Bai, *A Community-Based Routing Algorithm for Opportunistic Networks*, in International Conference on Ubiquitous and Future Networks. p. 701-706(2013).