

# Community Detection Using Maximum Connection Probability in Opportunistic Network

Yong Zhang, Ying Han  
Beijing Key Laboratory of Work  
Safety Intelligent Monitoring,  
Beijing University of Posts and  
Telecom, Beijing, P.R.C.  
Email: yongzhang@bupt.edu.cn,  
yingying88785@qq.com

Jin Li  
Beijing Information Science &  
Technology University  
Beijing, P.R.C.  
Email: bjlijin@126.com

Ping Wu  
Department of Engineering Sciences  
Uppsala University  
Uppsala, Sweden  
Email: pw@angstrom.uu.se

**Abstract**—A novel approach is proposed in this paper to detect community structure in opportunistic networks. Different from the existing solutions, this approach uses Maximum Connection Probability (MCP) instead of encounter probability. This approach is established in two phases. Firstly, an algorithm is proposed to derive the MCP of any node to other nodes. Secondly, the community structure derived from the MCP is identified using a divisive algorithm. Simulation is conducted based on walking day movement model to evaluate the approach. The results show that the proposed approach can detect community structure more accurately and reflect human relationship in reality.

**Keywords**- Community Detection; Opportunistic Networks; Connection Probability; Self-Organizing Networks; Network Structure

## I. INTRODUCTION

Opportunistic Networks (ONs) are dynamic, self-organizing networks which only can keep intermittent connectivity in the entire network [1]. It can be seen as the evolution of Mobile Ad hoc NETWORKS (MANETs). Mobile nodes with short-range radio communication capability in ONs cannot maintain the connectivity to other nodes even using multi-hop transmission. To communicate with mobile nodes in the disconnected part of the network, messages are stored in some nodes, carried to other places and forwarded to their neighborhood. The neighborhood also repeats the store-carry-forward process until the messages are received by the destination.

Mobility modeling is one of the most essential issues in ONs. The characteristic of human mobility model heavily depends on the relationships among the people carrying mobile nodes [2]. Because human movement has regular patterns in temporal and spatial scale, mobile nodes in ONs can be organized into different communities [3]. Grouping nodes into communities is helpful to highlight communication patterns and network characteristic. A lot of research works, such as routing and message forwarding

technology [4, 5], are carried out on the basis of the achievement on community detection.

Some researchers propose their methods to detect the community in ONs [3, 6, 7]. The theoretical foundation of most contributions is constructed on complex network science [8, 9, 10, 11]. Different evaluation metrics and grouping methods are investigated in existing literatures [3, 6, 7, 12, 13].

The authors in [6] analyze the community structure in Delay Tolerant Networks (DTNs) using three community detection algorithms (i.e., SIMPLE, k-CLIQUE, and MODULARITY) which are often used in complex networks science [10]. Eleonora B [3] proposes an improved algorithm named Adaptive Detection SIMPLE based on [4]. In these methods, the contact duration is adopted to represent the strength of relationship among nodes.

Ref. [7] points out the drawback of the method in [3] and proposes another community detection algorithm in opportunistic networks. The authors formulate message forwarding in opportunistic networks on finite graphs and analyze the algorithm performance based on the mathematic features of random walk mobility model. As we know, random walk model does not represent human movement characteristic well [14].

The contact duration or inter-contact time of one hop is not appropriate to represent the relationship strength among nodes. The communication occurs from end to end and the path from the source to the destination usually has more than one hop. In the existing literatures few works are seen on characterizing realistic behavior about end to end transmission. In this paper we are going to discuss the impact of relationship strength and propose a novel evaluation criterion instead of contact duration and inter-contact time, to evaluate the relationship strength. In particular, we adopt the connection probability from end to end as the evaluation criterion. We evaluate our proposed approach in a more realistic mobility mode, working day movement model [15, 16]. Simulation is going to be made

to show how well our approach works and how accurately it can detect the community.

The rest of this paper is structured as follows. We discuss existing measure criteria in community detection solutions and point out their drawback in Section 2. And then a novel approach named Maximum Connection Probability Detection approach (MCPD) is proposed in Section 3. We evaluate our proposal in ONE simulation platform in Section 4. Finally we conclude the paper with a brief discussion.

## II. PROBLEM FORMULATION

Currently, relationship among nodes is characterized by inter-contact time and contact duration. Inter-contact time is defined in the existing literatures [3, 6, 7, 16], as the time interval from the end of the contact to the beginning of next one. Contact duration is defined as the time interval during which two nodes are in radio range of each another. However, communication in ONs is a type of end to end behavior. One-hop transmission is not equal to the successful communication between source to destination. Most messages may be dropped by intermediate nodes because no path can be found by intermediate nodes before the time limit or other reasons. Communication capability from source to destination should be taken account of in community structure detection. Therefore, the relationship among nodes should be characterized by connection capability from end to end.

### A. Definitions

Before stating the problem, we define some other terminologies:

*Ties*: The tie means the link between any pair of nodes. It can be characterized by inter-contact time and contact duration. The two parameters can be transformed as the weights of interactions between mobile nodes and express how frequently and for how long two nodes spend time together [13, 16, 17]. It also can be characterized by Encounter Probability (EP) which is the ratio of the contact duration to the whole time.

*Path*: The path includes one or several links through which the messages can be transmitted from source to destination.

*Connection Probability (CP)*: CP of one path in this paper especially specifies the connection probability from source to destination. It can be calculated by the product of the EPs of all the ties along this path.

### B. Tie strength and its impact

In this paper a tie with long contact duration or a large EP is called strong tie. In contrast, a weak tie has short contact duration or a low EP. The weight of a tie is expressed by EP. The weak ties whose weights are lower than a threshold are ignored in many research works [3, 7]. Especially, in some routing protocols such as HiBOP[18], HCR[19], etc., the messages are forwarded only on the strong ties. As we know, the weak ties may play different roles in networks [20] and should not be omitted in transmission process. The roles of

strong ties and weak ties are illustrated in Figure 1. Node A transmits message to node B.

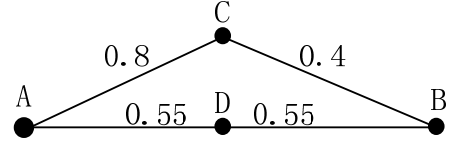


Figure 1. Figure 1 Sample for ties strength

Consider the threshold  $w^* = 0.5$ . If the link has a weight of  $w$  satisfying  $w < w^*$ , it is a weak tie, and the transmission does not occur in this link. In Figure 1, there are two paths from node A to node B. Path ‘A-D-B’ has the CP  $0.55 \times 0.55 = 0.3025$ , and path ‘A-C-B’ has  $0.4 \times 0.8 = 0.32$ . This shows that the path ‘A-C-B’ has MCP. And the total connection probability through the two disjoint paths is  $1 - (1 - 0.32)(1 - 0.3025) = 0.5257$ . From this example we can find that the ties strength cannot measure the relationship among nodes directly. It is more accurate that the evaluation criterion should adopt the CP than EP.

However it can be very difficult to calculate CPs when there are a lot of paths from a source to a destination. When these paths may own the same ties and correlate each other, exactly calculating CPs in ONs is unpractical. And the path with MCP is unique and chosen as the transmission path. Based on this hypothesis, we propose a novel approach named Maximum Connection Probability Detection (MCPD) to find the path with MCP among all mobile nodes.

## III. MAXIMUM CONNECTION PROBABILITY DETECTION APPROACH

Because the human movement has regular patterns in time and space scales [2, 16], we investigate the node’s behavior in terms of days. Assume the network runs for some days and each node can record its EP with other nodes. Our approach is established in two phases. Firstly, we derive the MCP of any node to other nodes. Secondly, we detect the community structure from MCP.

Consider an undirected connected graph  $G = (V, E)$ , where  $V$  is the set of nodes and  $E$  is the set of edges. An edge means a tie between a pair of nodes. Let  $|V| = n$  and  $|E| = m$ . Each edge  $(u, v)$  is assigned a weight  $p_{u,v}$  ( $u, v \in V, (u, v) \in E$ ) to represent EP. Let  $p_{u,v} = p_{v,u}$  due to the undirected graph. The weight graph can be represented by adjacency matrix where

$$w_{u,v} = \begin{cases} p_{u,v}, & \text{if } (u, v) \in E \\ \infty, & \text{if } (u, v) \notin E \\ 0, & \text{if } i = j \end{cases}$$

The first phase of the MCPD is to find MCP. Herein, the approach runs just like Floyd algorithm which is to find the shortest path in a graph. Unlike the Floyd algorithm, our goal

is to find MCP. If the path from source node  $u_0$  to destination node  $u_k$  includes edges set  $\bigcup_{i=0}^{k-1} (u_i, u_{i+1})$ , the CP of this path is

$$P_{u_0, u_k} = \prod_{i=0}^{k-1} w_{u_i, u_{i+1}}$$

The approach in phase 1 runs as follows:

Step 1.1. Initialize the probability matrix  $W^{(0)}$  and routing matrix  $R^{(0)}$  as

$$W^{(i)} = \begin{pmatrix} w_{1,1}^{(i)} & \cdots & w_{1,v}^{(i)} & \cdots & w_{1,n}^{(i)} \\ \vdots & & \vdots & & \vdots \\ w_{u,1}^{(i)} & \cdots & w_{u,v}^{(i)} & \cdots & w_{u,n}^{(i)} \\ \vdots & & \vdots & & \vdots \\ w_{n,1}^{(i)} & \cdots & w_{n,v}^{(i)} & \cdots & w_{n,n}^{(i)} \end{pmatrix}, 0 \leq i \leq n,$$

$$w_{u,v}^{(0)} = \begin{cases} p_{u,v}, & \text{if } (u,v) \in E \\ \infty, & \text{if } (u,v) \notin E \\ 0, & \text{if } i = j \end{cases}$$

$$R^{(i)} = \begin{pmatrix} r_{1,1}^{(i)} & \cdots & r_{1,v}^{(i)} & \cdots & r_{1,n}^{(i)} \\ \vdots & & \vdots & & \vdots \\ r_{u,1}^{(i)} & \cdots & r_{u,v}^{(i)} & \cdots & r_{u,n}^{(i)} \\ \vdots & & \vdots & & \vdots \\ r_{n,1}^{(i)} & \cdots & r_{n,v}^{(i)} & \cdots & r_{n,n}^{(i)} \end{pmatrix}, 0 \leq i \leq n,$$

$$r_{u,v}^{(0)} = \begin{cases} v, & \text{if } w_{u,v}^{(0)} \neq \infty \\ 0, & \text{others} \end{cases}$$

Step 1.2. Find the  $w_{u,v}^{(i)}$  and  $r_{u,v}^{(i)}$  recursively, in the following manner

$$w_{u,v}^{(i)} = \max(w_{u,v}^{(i-1)}, w_{u,i}^{(i-1)} \times w_{i,v}^{(i-1)})$$

$$r_{u,v}^{(i)} = \begin{cases} r_{u,i}^{(i-1)} & \text{if } w_{u,v}^{(i)} > w_{u,v}^{(i-1)} \\ r_{u,v}^{(i-1)} & \text{if } w_{u,v}^{(i)} = w_{u,v}^{(i-1)} \end{cases}$$

Step 1.3. If  $i < n$ ,  $i = i+1$  and goto step 1.2. Otherwise, algorithm stops if  $i = n$ .

Having finished the calculation in this phase, we find the MCP as  $C_{u,v} = w_{u,v}^{(n)}$

The routing matrix  $R^{(n)}$  indicates the path with MCP from node  $u$  to node  $v$ .  $C_{u,v}$  represents a new relationship graph among nodes whose edge's weight represents the MCP from node  $u$  to node  $v$ . In the next phase, we adopt divisive algorithm to detect the community structure [10]. The difference between our approach and traditional divisive algorithm proposed by Newman [10, 11] is that we use the MCP instead of edge betweenness. A threshold weight  $w^*$  ( $0 \leq w^* < 1$ ) is set up to measure whether an edge should be removed.

The approach in phase 2 runs as follows:

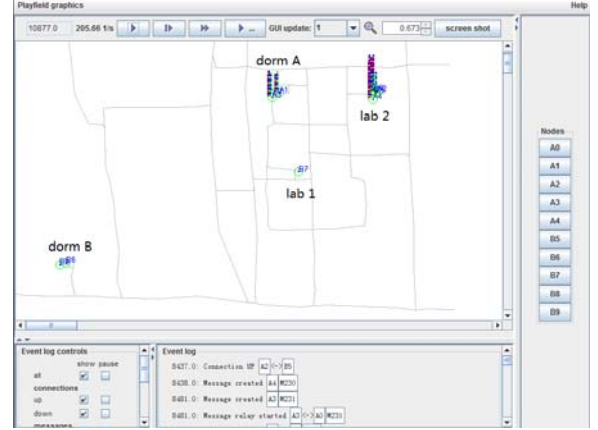


Figure 2. Simulation Scenario

Step 2.1. Initially, this threshold weight is assigned a small value, for example  $w^* = 0.1$ . Initialize any  $\varepsilon$  ( $0 < \varepsilon < 1$ ) to represent an incremental value.

Step 2.2. The edge  $(u,v)$  is removed if  $C_{u,v} < w^*$ .

Step 2.3. Examine whether the communities take shape. If not, increase  $w^*$  by  $w^* = w^* + \varepsilon$  and goto step 2.2..

#### IV. SIMULATION AND EVALUATION

In this section, we evaluate our approach using MCP matrix in comparison with EP matrix. We use ONE [21] to simulate human behavior and collect EP. The experiment scenario is shown in Figure 2 which is the map of Beijing University of Posts and Telecommunications with the size of roughly  $1000 \times 1000 m^2$ . The mobility model is Working Day Movement model [15]. Based on the life habits of ten students in our research team, we collect their information about dormitory, laboratory, friends and interest points in two days. They work in office, stay in dormitory or take parties in the evening activity. The ten students are divided into two groups. Group A including  $\{0, 1, 2, 3, 4\}$  lives in dormitory A. Group B including  $\{5, 6, 7, 8, 9\}$  lives in dormitory B. Node 0, 1 and 7 work in lab 1 and other members work in lab 2. They work 8 hours in a day, and may wander along the road in the map in campus after work. The probability of doing some evening activities after work was set to 0.5. The office size is  $100 \times 100 m^2$ . The transmit range of nodes is 10 meters. It means that the connection is established if two nodes within the range of 10 meters. The other settings are similar to those in [15].

These parameters are input to ONE simulation platform and collect the datasets in two days. So we get the EP matrix as:

$$E_p = W^{(0)}$$

1.0000	0.2503	0.2756	0.7650	0.5884	0.0001	0	0.0116	0	0
0.2503	1.0000	0.0705	0.2637	0.2786	0	0	0.0067	0	0
0.2756	0.0705	1.0000	0.0002	0.0085	0.1237	0	0	0.1443	0
0.7650	0.2637	0.0002	1.0000	0.3483	0.0838	0.1356	0	0	0.0678
0.5884	0.2786	0.0085	0.3483	1.0000	0.0521	0.0375	0	0.0046	0.0632
0.0001	0	0.1237	0.0838	0.0521	1.0000	0.0708	0.0001	0.1425	0.3909
0	0	0	0.1356	0.0375	0.0708	1.0000	0.0004	0.0003	0.0910
0.0116	0.0067	0	0	0	0.0001	0.0004	1.0000	0.0510	0.2545
0	0	0.1443	0	0.0046	0.1425	0.0003	0.0510	1.0000	0.2523
0	0	0	0.0678	0.0632	0.3909	0.0910	0.2545	0.2523	1.0000

The CP matrix is

$$C_p = W^{(10)}$$

1.0000	0.2503	0.2756	0.7650	0.5884	0.0641	0.1037	0.0132	0.0398	0.0518
0.2503	1.0000	0.0705	0.2637	0.2786	0.0221	0.0358	0.0067	0.0102	0.0179
0.2756	0.0705	1.0000	0.2108	0.1621	0.1237	0.0286	0.0123	0.1443	0.0484
0.7650	0.2637	0.2108	1.0000	0.4501	0.0838	0.1356	0.0172	0.0304	0.0678
0.5884	0.2786	0.1621	0.4501	1.0000	0.0521	0.0610	0.0161	0.0234	0.0632
0.0641	0.0221	0.1237	0.0838	0.0521	1.0000	0.0708	0.0995	0.1425	0.3909
0.1037	0.0358	0.0286	0.1356	0.0610	0.0708	1.0000	0.0232	0.0230	0.0910
0.0132	0.0067	0.0123	0.0172	0.0161	0.0995	0.0232	1.0000	0.0642	0.2545
0.0398	0.0102	0.1443	0.0304	0.0234	0.1425	0.0230	0.0642	1.0000	0.2523
0.0518	0.0179	0.0484	0.0678	0.0632	0.3909	0.0910	0.2545	0.2523	1.0000

The routing matrix is

$$R^{(10)} = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 3 & 3 & 3 & 2 & 3 \\ 0 & 1 & 2 & 3 & 4 & 3 & 3 & 7 & 2 & 3 \\ 0 & 1 & 2 & 0 & 0 & 5 & 0 & 5 & 8 & 5 \\ 0 & 1 & 0 & 3 & 0 & 5 & 6 & 9 & 0 & 9 \\ 0 & 1 & 0 & 0 & 4 & 5 & 0 & 9 & 0 & 9 \\ 3 & 3 & 2 & 3 & 4 & 5 & 6 & 9 & 8 & 9 \\ 3 & 3 & 3 & 3 & 3 & 5 & 6 & 9 & 9 & 9 \\ 9 & 1 & 9 & 9 & 9 & 9 & 9 & 7 & 9 & 9 \\ 2 & 2 & 2 & 2 & 2 & 5 & 9 & 9 & 8 & 9 \\ 3 & 3 & 5 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{bmatrix}$$

We draw the relationship graphs in terms of EP matrix and MCP matrix (Figure 3). The community structure is identified and shown in Figure 4 while we remove all the edges whose weights are lower than the threshold  $w^*$  ( $w^* = 0.1$ ). The community structure in Figure 4(b) is more distinguished. Node 2 in Figure 4(b) links the community  $\{0, 1, 3, 4, 6\}$  and community  $\{5, 7, 8, 9\}$ . We observe the nodes' behavior in the simulation process and then find that node 2 wanders for a long time after work. Node 0 and node 2 take party in the evening of the second day. From the information, we can find that node 2 is an

active node which may be a mobile infrastructure node. This is an important attribution which can impact on data forwarding in ONs [22]. But we cannot find the information in figure 4(a).

As is shown in Figure 5, the two graphs can be divided into three parts, ie., community  $\{0, 1, 2, 3, 4\}$ , community  $\{5, 7, 8, 9\}$  and isolated node 6 while we increase threshold  $w^*$  to 0.2. The difference between Figure 5(a) and Figure 5(b) is the relationship between node 2 and node 3. The encounter probability between node 2 and node 3 is 0.0002 due to EP matrix  $E$ . But from routing matrix  $R^{(10)}$ , we can find the path from node 2 to node 3 is '2-0-3'. Although the direct tie between node 2 and node 3 is weak, the MCP is higher. This reveals that the MCP matrix is more accurate to display the nodes' relationship than EP matrix.

Furthermore, we define the node's weight as the sum of the links weights which are associated with this node. This index can represent node's activity in community. In Figure 5(a) and 5(b), the node with maximum weight is node 0. But the node with the second largest weight is node 3 in Figure 5(b) instead of node 4 in Figure 5(a). It means the messages which are forwarded by node 3 can arrive at other nodes with larger probability than node 4. Also Figure 5(a) cannot express this information.

## V. CONCLUSION

Community detection technology is a crucial issue for the development of ONs. Currently, all the community detection solutions use contact duration or encounter probability as evaluation criteria. After investigating these approaches, we redefine the relationship among nodes and propose a new approach named MCPD based on MCP instead of EP. We evaluate the MCPD in a realistic mobile model and find that the MCPD can detect the community structure more accurately. In the new community structure, active nodes which link several communities can be detected even though the active nodes have small contact duration with other nodes. And the nodes with large weight can be distinguished using MCPD. Some routing protocols based on mobile infrastructure can be implemented before these nodes in community are detected correctly. From this perspective, MCPD lays a sound foundation for these researches based on community structure in ONs.

## ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (NO. 61171097).

## REFERENCES

- [1] Stavroulaki V, Tsagkaris K, Logothetis M, et al. Opportunistic networks. Vehicular Technology Magazine, IEEE, 2011, 6(3): 52-59
- [2] Musolesi M, Mascolo C. A community based mobility model for ad hoc network research. In: Proc. of the 2nd International Workshop on Multi-hop Ad Hoc Networks: from Theory to Reality, New York: ACM, 2006. 31-38
- [3] Eleonora B, Marco C, Andrea P. Autonomic detection of dynamic social communities in Opportunistic Networks. Ad Hoc Networking

- Workshop (Med-Hoc-Net), 2011 The 10th IFIP Annual Mediterranean, Favignana Island, Sicily, Italy, 2011. 142-149
- [4] Niu J W, Zhou X, Wang K Q, et al. A data transmission scheme for community-based Opportunistic Networks. 5th International Conference on Wireless Communications, Networking and Mobile Computing, Beijing, China, 2009. 1-5
  - [5] Hui P, Crowcroft J, Yoneki E. BUBBLE Rap: Social-based forwarding in Delay Tolerant Networks. IEEE Transactions on Mobile Computing, 2011, 10(11):1576-1589
  - [6] Hui P, Yoneki E, Chan S-Y, et al. Distributed community detection in delay tolerant networks. In: Proceedings of MobiArch'07, Kyoto, Japan, 2007. 1-8
  - [7] Xu K, Yang G H, Li V, et al. Detecting dynamic communities in Opportunistic Networks. In: Proc. of First International Conference on Ubiquitous and Future Networks, Hong Kong, China, 2009.159-164
  - [8] Newman M E J. Detecting community structure in networks. The European Physical Journal B, 2004, 38(2):321-330
  - [9] Palla G, Derenyi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society, Nature,2005, 435(9):814-818
  - [10] Fortunato S. Community detection in graphs. Physics Reports, 2010, 486:75-174
  - [11] Newman M E J, Girvan M. Finding and evaluating community structure in networks. Physical Review. E,2004, 69(2):15
  - [12] Katsaros D, Dimokas N, Tassioulas L. Social network analysis concepts in the design of wireless Ad Hoc Network protocols. IEEE Network, 2010, 24(6): 23-29
  - [13] Karagiannis T, Boudec L, Vojnović M. Power Law and exponential decay of intercontact times between mobile devices. IEEE Transactions on Mobile Computing, 2010, 9(10):1377-1390
  - [14] Camp T, Boleng J, Davies V. A survey of mobility models for Ad Hoc Network research. Wireless Communications and Mobile Computing, Wiley,2002, 2(8): 483-502
  - [15] Ekman F, Keranen A, Karvo J, et al. Working day movement model. Proceedings of the 1st ACM SIGMOBILE workshop on Mobility model, 2008.33-40
  - [16] Karamshuk D, Boldrini C, Conti M, et al. Human mobility models for Opportunistic Networks. IEEE Communications Magazine, 2011, 49(12): 157-165
  - [17] Chaintreau A, Hui P, Crowcroft J, et al. Impact of human mobility on opportunistic forwarding algorithms. IEEE Transactions on Mobile Computing, 2007,6(6): 606-620
  - [18] Boldrini C, Conti M. HiBOP: a history based routing protocol for opportunistic networks. 2007 IEEE International Symposium on a World of wireless, Mobile and Multimedia Networks, Espoo, Finland, 2007. 1-12
  - [19] Xie X G, Zhang Y, Dai C, et al. Heterogeneous context-aware routing protocol for Opportunistic Network.2010 5th International Conference on Pervasive Computing and Applications, Maribor, Slovenia, 2010. 238-243
  - [20] Granovetter M S. The strength of weak ties. The American Journal of Sociology, 1973, 78(6):1360-1380
  - [21] Keronen A, Ott J. The ONE simulator for DTN protocol evaluation. Proceedings of the 2nd International Conference on Simulation Tools and Techniques, Rome, Italy, 2009. 1-10
  - [22] Pelusi L, Passarella A, Conti, M. Opportunistic Networking: Data Forwarding in Disconnected Mobile Ad Hoc Networks, IEEE Communications Magazine, 2006, 44(11): 134-141

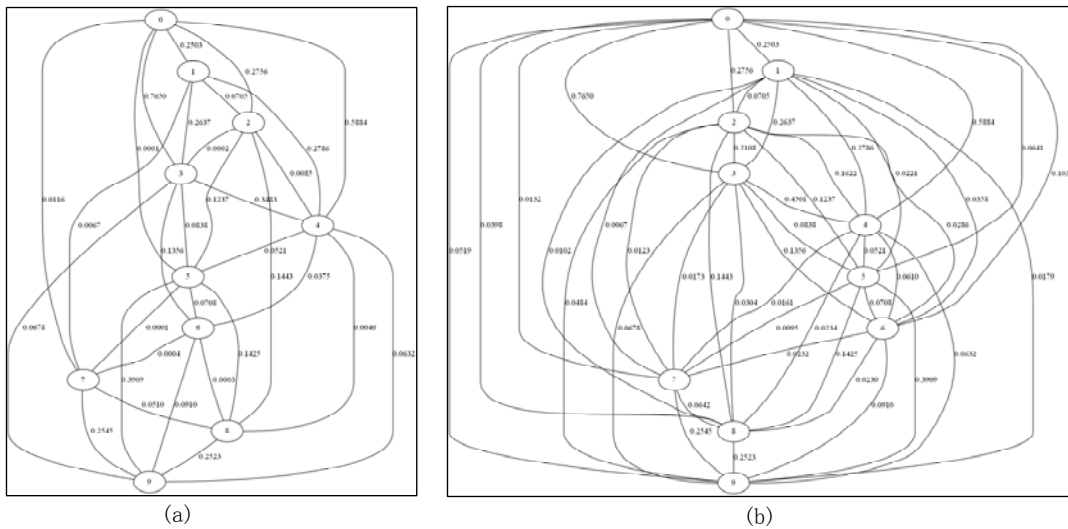


Figure 3. Relationship graph. ((a) Based on EP matrix; (b) based on MCP matrix)

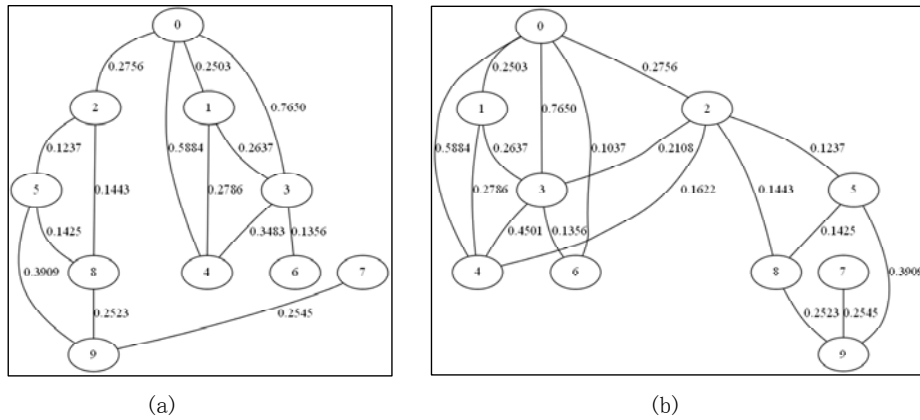


Figure 4. Relationship graphs (Note that all edges whose weights are lower than 0.1 are removed) ((a) Based on EP matrix; (b) based on MCP matrix)

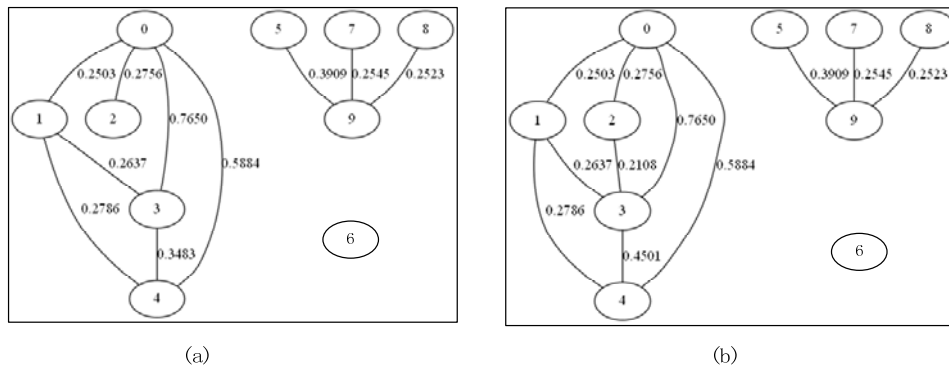


Figure 5. Relationship graph (Remove all edges whose weights are lower than 0.2) ((a) Based on EP matrix; (b) based on MCP matrix)