

# A Study on the Node Centrality Based Multi-social Attributes Weighted in Mobile Social Networks

Yanhong Meng<sup>1,\*</sup>, Xianxian Liu<sup>1</sup>, Peirong Zhao<sup>1</sup>, and Yunhui Yi<sup>1</sup>

<sup>1</sup>State Key La Laboratory of Integrated Services Networks, Xidian University, Xi'an, China

**Abstract.** Mobile social networks (MSNs) exploit human mobility and consequent device-to-device contact to opportunistically realize data communication. Thus links in MSNs is dynamic changing over time and strongly influenced by people activities, mining influential nodes is one of the important questions for effective information transmission in MSNs. While traditional centrality definitions are based on the static binary network model and not suitable for time-varying topology structure in mobile social network. Furthermore previous centrality metrics often referred to social attributes about neighbor nodes and contact times, and did not take the contact duration time into consideration. Therefore, this paper proposes a centrality measurement method based on multi-social attributes weighted. We first use the temporal evolution graph model which more accurately depicts the dynamic nature of topology in MSNs. Quantifying human social relations and mobility model as weights for the links, and then we redefine degree of centrality and the measurement of shortest path. Finally, the superiority of the concepts we posed are evaluated in the real data set.

## 1 Introduction

The rapid development of Internet, the advent of the era of 4G mobile Internets, the fast growth of mobile terminals technology and the massive popularity of mobile terminals prompt the mobile social networks emerging at the historic moment. MSNs can take advantage of mobile features of wireless devices users and ubiquitous accessibility, and finally realize information transmit among interested mobile users without relying on mobile networks operators' communication infrastructures [1-3, 10].

In mobile social network, node is represented as people and group, meanwhile relation shows the relationship between nodes [4-5]. The problem is how to find the most influential node in the network. The number of important nodes is generally very small, but their impact can more quickly spread to most other nodes in the network. Many scholars use the central metric method to measure the importance of the nodes. The most typical one is proposed by Freeman [17]. While traditional centrality definitions are based on the static binary network model and not suitable for time-varying topology structure in MSNs. So

---

\* Corresponding author: [yhmeng@stu.xidian.edu.cn](mailto:yhmeng@stu.xidian.edu.cn)

some researchers extend centralities measures to weighted networks [6, 7] and dynamic networks[9-11, 16]. Opsahl et al. [7] generalizes Freeman’s three centralities, not only focusing on tie weights, but also on the number of ties. Temporal network can be modelled in terms of time-varying graphs, which are time-ordered sequences of graphs over a set of nodes, centralities measurement must obey the time [9-11].

The sum of various human relations form social attributes, what’s more, links in MSNs is dynamic changing over time and strongly influenced by people’s activities. Paper [16] puts forward entropy degree centrality (EDC) and social-based global centrality that both based on social relations (neighbor nodes and contact times), when contact time equals 1, no matter how many neighbor numbers are, we get the same EDC. While common neighbor is zero, no matter how big the contact time is, the social-based global centrality always is same, thus ignore impact of the different contact times. According to the above discussions, in this paper, we add contact duration time and propose a centrality measurement method based on multi-social attributes weighted.

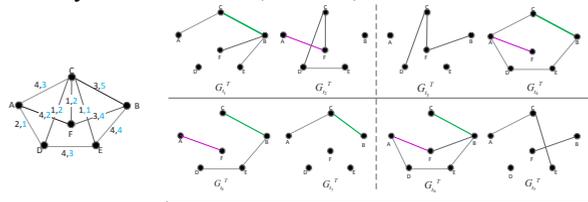
The rest of this paper is organized as follows: Sec.2 gives a brief overview of preliminary works. Measurement of centrality based multiple social attributes in MSNs are described in Sec.3. In Sec. 4, we present the experiment results of all proposed concepts. Finally, we offer conclusions in Sec. 5.

## 2 Preliminaries

### 2.1 Temporal evolution graph model

We use temporal evolution graph model [9-11, 16] to provide a concise and general formulation that takes into account time order, contact duration times and contact times to capture the dynamic topology of MSNs. The relations between entities are assumed to take place over a time span  $T$ . We use  $\omega$  to denote the time duration of each snapshot (or time window size).  $Q^T = T/\omega$ , expressed numbers of time units. So dynamic networks will be represented as  $G^T = (V^T, E^T, M^T)$ ,  $G^T = \{G_{t_1}^T, G_{t_2}^T, \dots, G_{t_{Q^T}}^T\}$  during the duration time  $T$ .  $V^T: V^T \subseteq V$ , the set of all individuals in the network during  $T$ .  $E^T: E^T \subseteq E$ , the set of edges that represent contact events between individuals during  $T$ .  $M^T$ : a sequence of connectivity matrices which record contact events of between individuals during  $T$ .

We introduce the following illustrative example in Fig.1, the left graph is the total static graph in time  $T$ , the connections between two points indicate the contact times (black number) and the total contact duration time (blue number). The right graph depicts dynamic connect graphs in different time units. We assume  $T = 8$  mins ,  $\omega = 1$ min and  $Q^T = 8$ . Observing edge A-F, there are only four contact times in time units 2, 4, 6, and 8. While contact duration time is 2mins, so the half time must be invalid in four times connection. However, due to the window size is bigger this fact cannot be observed precisely. Then analysing edge B-C, we can see that it exists 3 times connect in 1, 4, and 6/7/8 time units, because B and C are always connected in 6, 7 and 8, we consider it as one time connect.



**Fig. 1.** A simple example about temporal evolution graph

## 2.2 Measurement of centrality in opsahl weighted network

### 2.2.1 Degree centrality

Freeman [17] asserted that the degree of a focal node is the number of adjacencies in a binary network. For one node, the greater the number of neighbors are, the stronger the connect ability to other nodes. So we give the definition  $k_i = C_D(i) = \sum_j^N x_{ij}$ , where  $i$  is the focal node,  $j$  represents all other nodes,  $N$  is the total number of nodes, and  $x$  is adjacency matrix, in which  $x_{ij}$  denotes 1 if node  $i$  is connected to  $j$ , and 0 otherwise.

Degree has generally been extended to the sum of weights when analyzing weighted network, and labeled node strength, the definition is  $s_i = C_D^w(i) = \sum_j^N w_{ij}$ , where  $w$  is the weighted adjacency matrix, in which  $w_{ij}$  is greater than 0 if the node  $i$  is connected to node  $j$ , and the value represents the weight of the tie.

However, measurement of degree in binary network and weighted network focuses on either tie weights or the number of ties, and they are both incomplete measurement. In an attempt to combine both degree and strength, Opsahl et al.[7] import a tuning parameter,  $\alpha$ , which determines the relative importance of the number of ties compared to ties weights. The formula is as follows:

$$C_D^{w\alpha}(i) = k_i \times \left(\frac{s_i}{k_i}\right)^\alpha = k_i^{1-\alpha} \times s_i^\alpha \quad (1)$$

Where  $\alpha$  is a positive tuning parameter, if this parameter is between 0 and 1, then having a high degree is taken as favourable.

### 2.2.2 Closeness centrality and betweenness centrality

The closeness and betweenness centrality measures rely on the identification and length of the shortest paths among nodes in the network. In binary network, Freeman [17] defined the shortest path formulation as  $d(i, j) = \min(x_{ih} + \dots + x_{hj})$ , from which we can observe the shortest path in binary network can only impacted by the number of relay nodes passed by.

In weighted network, there were several expressive methods, Dijkstra [12] proposed an algorithm that finds the path of least resistance, and was defined for networks where the weights represented costs of transmitting. Newman et al[13] and Brande et al.[14] defined Dijkstra algorithm as  $d^w(i, j) = \min(1/w_{ih} + \dots + 1/w_{hj})$ , from which we can also find the shortest path in weighted network can only impacted by the weights between nodes.

In order to focus both on the number of relay nodes and the weights between nodes, Opsahl et al.[7] also import a tuning parameter,  $\alpha$ , and redefine the shortest path as follows:

$$d^{w\alpha}(i, j) = \min\left(\frac{1}{(w_{ih})^\alpha} + \dots + \frac{1}{(w_{hj})^\alpha}\right) \quad (2)$$

The impact of different tuning parameters will be analyzed, when  $\alpha = 0$ , the proposed measure produces the same outcome as the binary distance measure, whereas when  $\alpha = 1$ , the outcome is the same as the one obtained with Dijkstra's algorithm. A value for

$0 < \alpha < 1$  assigns the path with the greatest number of intermediary nodes has the longest distance.

### 3 Measurement of centrality based multiple social attributes in MSNs

In mobile social networks, the sum of various human relations form social attributes, so the human mobility bring fierce impact. Paper [8] suggested that the strength of a relationship is dependent on four components: the frequency of contact, the length or history of the relationship, contact duration, and the number of transactions. In this paper, we take contact times, contact duration time, the number of neighborhood node and common neighborhood node into consideration to quantify the weight between nodes.

#### 3.1 Dissimilar weight and similar weight

Literature [13] lists weight measurement in the social network and non-social network, generally speaking, a social network, the connection weights are a fuction of duration, emotional intensity, intimacy and the exchange of service.

According to different meaning of weight, weights can be divided into two categories: dissimilar weight and similar weight. In general, the most we have studied is the dissimilar weight which corresponding to the distance and the larger the weight the further the distance between two nodes. However, the similarity weight stands the opposite. In fact, similar weight is prevalent in the social network, for example, in the cooperative process of scientists, the weight is higher on behalf of the cooperation between the two scientists is more frequently and more closely. So we can assume that the smaller the distance between them.

The differences between dissimilar weight and similar weight directly affect the definition of shortest path length of network. Given a path from node  $i$  to  $j$  and pass by node  $k$ , for dissimilar weight, and the distance between node  $i$  and node  $j$  is  $d_{ij} = w_{ik} + w_{kj}$ , while the distance is defined  $d_{ij} = 1/w_{ik} + 1/w_{kj}$  for similar weight.

#### 3.2 Determination of tie strength (or weight)

Due to similar weight is prevalent in the social network, in this paper, we take contact times, contact duration time, the number of neighborhood nodes and common neighborhood nodes into consideration to quantify the tie strength (or weight) between nodes. Tie strength between nodes is determined by multiple social attributes. Give a time period  $T$ , we assume that there are  $M$  factors,  $factor_k, k = 1, \dots, M$  and  $g_i(factor_k)$  means the contribution to tie strength between node  $i$  and  $j$  for any factor. Then the contributions of  $M$  factors are as follows:  $TieStrength^T(i, j) = \{g_1(factor_1), \dots, g_M(factor_M)\}$ .

Basagni et al. [15] shows the high correlation between contact times (regularity) and contact duration (familiarity) in four datasets, and they define four kinds of relationships between a pair of nodes: community, familiar strangers, strangers, and friends. For a pair of nodes, the greater the encounter frequency, the greater the chance to communicate. If two nodes stay a longer time, the like is more stable. On other hands, if two nodes have more other acquaintances in common, they may have a higher probability of two people being acquainted. So we propose the tie strength, as:

$$TieStrength^T(i, j) = F(i, j)^{\delta+sim(i, j)} + D(i, j)^{\delta+sim(i, j)} \quad (3)$$

Where  $F(i, j)$  and  $D(i, j)$  represents contact times and contact duration time between node  $i$  and node  $j$  during the time span  $T$ , respectively.  $\delta$  is correction factor, whose value range is  $0 < \delta \ll sim$ , which makes the equ.(3) not equal 1, when their common nodes is 0, but contact times and duration time are not 0. Where  $N(i)$  is the neighborhood number,  $sim(i, j)$  stands for density of common nodes, and is denoted as follows:

$$sim(i, j) = \frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|} \tag{4}$$

### 3.3 Multi-social attributes degree centrality

According to the tie strength proposed in Sec.3.2, we get the node strength during a time span  $T$  as:

$$S^T_{iM-social} = \sum_{j \in N(i)} TieStrength^T(i, j) \tag{5}$$

Where  $N$  is the neighborhood of node  $i$ , but equ.(5) also does not consider neighborhood number, so from the discussion of Sec. 2.2.1, multi-social attributes degree centrality is denoted as:

$$C^T_{DM-social}^{w\alpha}(i) = k_i \times \left( \frac{S^T_{iM-social}}{k_i} \right)^\alpha = k_i^{1-\alpha} \times (S^T_{iM-social})^\alpha \tag{6}$$

We also create a tuning parameter  $\alpha$  as:

$$\alpha = \frac{\left( S^T_{1M-social} \times \dots \times S^T_{iM-social} \times \dots \times S^T_{NM-social} \right)^{\frac{1}{N}}}{\max \left\{ S^T_{1M-social}, \dots, S^T_{iM-social}, \dots, S^T_{NM-social} \right\}} \tag{7}$$

After analyzing the equ.(7), we know  $0 < \alpha < 1$ . Combing with equ.(1), under the same consideration, neighborhood number is much more, the degree centrality is bigger.

### 3.4 Multi-social attributes closeness centrality and betweenness centrality

From the discussion of Sec. 3.1, we see that the greater the weight, the smaller the distance between two nodes, and that is to say the tighter between two nodes the distance of two nodes in MSNs. In addition, when the information goes through more nodes, the more likely it is to be stolen or tampered. Thus less hops is hoped in MSNs.

Assuming a path from node  $i$  to node  $j$  will pass  $m$  nodes in a span time  $T$ ,  $n_1, n_2, \dots, n_m$ ,  $w_{ij} = TieStrength^T(i, j)$ , then the distance from node  $i$  to node  $j$  defined as:

$$d^w_{M-social}(i, j) = \min \left( \frac{1}{(w_{in_1}^T)^\alpha} + \frac{1}{(w_{n_1n_2}^T)^\alpha} + \dots + \frac{1}{(w_{n_mj}^T)^\alpha} \right) \tag{8}$$

Where a tuning parameter denoted as:

$$\alpha = \frac{\left( w_{in_1}^T \times w_{n_1n_2}^T \times \dots \times w_{n_mj}^T \right)^{\frac{1}{m+1}}}{\max \left\{ w_{in_1}^T, w_{n_1n_2}^T, \dots, w_{n_mj}^T \right\}} \tag{9}$$

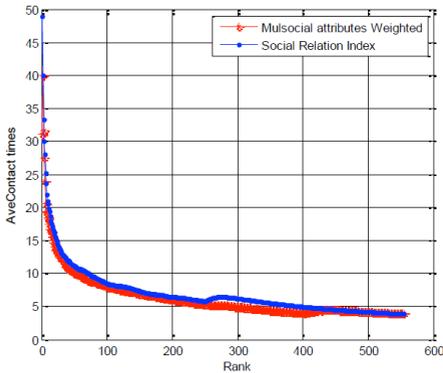
After analyzing the equ.(9), we know  $0 < \alpha < 1$ . Comparing with equ.(2), under the same consideration, a path of two nodes has less hops, the distance is smaller.

## 4 Experiment results and analysis

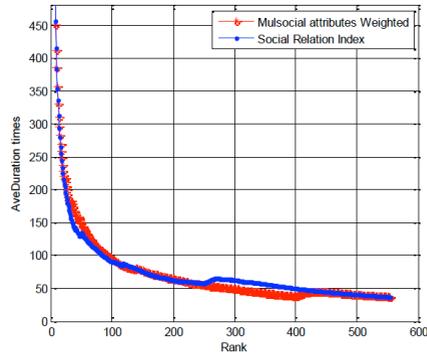
### 4.1 MIT reality data set

The Reality Mining project[18] was conducted from 2004–2005 at the MIT Media Laboratory. They used mobile phones pre-installed with several pieces of software to record the researchers' data. Subjects were observed using these measurements over the course of nine months and included students and faculty from two programs within a major research institution. The bluetooth scan interval was 5 min. We use the bluetooth traces over one month in 2004 October and define the contact as the link during which two devices are in mutual radio transmission range and can exchange data. If two scan time less than 30 min, we believe the two adjacent contact as one contact.

### 4.2 Experimental results of the tightness between nodes



**Fig. 2.** Average contact times in tightness Ranking top N.

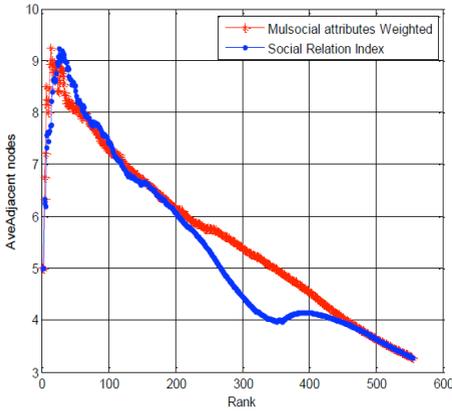


**Fig. 3.** Average duration time in tightness ranking top N.

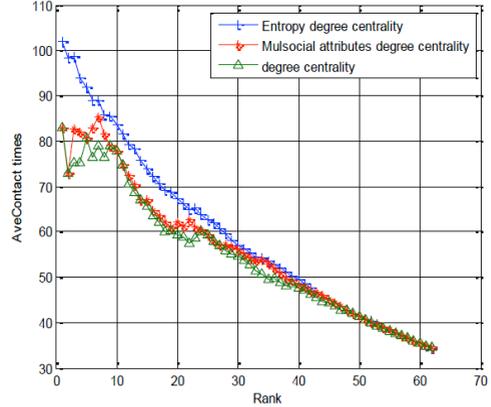
In order to evaluate the centrality based on multi-social attributes, we first compare the tie strength proposed in this paper with social relation index defined in paper [16] about average contact times, average contact duration time, and average common node number in tightness ranking top N. Figure 2, 3 and 4 report the tie experiment results between strength and SRI by MIT traces, and horizontal axes are the tightness ranking top N in three figures, besides vertical axes represent average contact times, average contact duration time, and average common node number in three figures, respectively. Because contact duration is a new factor added in tie strength and its constraint to contact times, figure 2 shows that the average contact times in SRI is slightly high compared with tie strength. Conversely, figure 3 shows that the average duration times in tie strength is slightly high compared with SRI before rank 251. In Fig. 4, it shows most rank tie strength is higher, while only rank between 18 and 108, the average common node number in SRI is slightly high compared with tie strength. Because when common neighbor is 0, no matter how big the contact times and duration are, SRI equal 1, which leads to ignoring the influences of different contact times. This is why the curve about SRI in figure 2 rises abruptly after rank 251, and also goes up suddenly in figure 3. A sharp decline appears in figure 4 because of

common nodes become 0. Above all, measurement of tightness proposed in this paper can considers more comprehensive factors and get more smooth curse compared with SRI.

### 4.3 Experimental results of multi-social attributes degree centrality (MSADC)

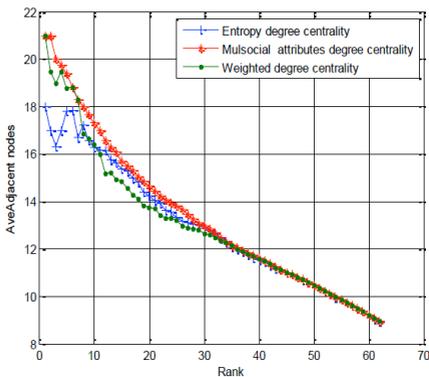


**Fig. 4.** Average adjacent nodes in tightness ranking top N.

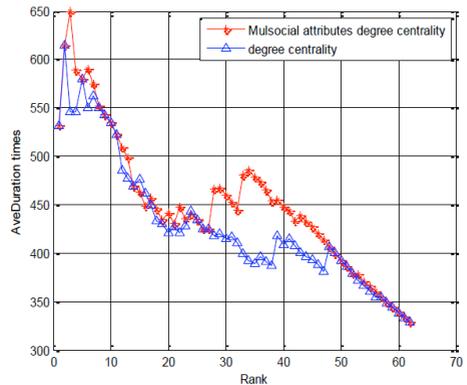


**Fig. 5.** Average contact times in MSADC ranking top N.

We compare MSADC with entropy degree centrality[16], degree centrality[17], and weighted degree centrality[14]. Figure 5 shows the the curves of average number of contacts of the top-N nodes which are ranked by MSADC, entropy degree centrality, and weighted degree centrality. Owing to duration time the new factor added in MSADC and its constraint to contact times, the curse about MSADC is lower than entropy degree centrality and higher than degree centrality. But the average number of contacts of the top-N nodes about MSADC is most higher than others in Fig. 6. For not thinking over the duration time in entropy degree centrality, figure 7 just reports the the curves of average duration time of the top-N nodes which are ranked by MSADC and weighted degree centrality. We can learn that MSADC precedes degree centrality in the aspect of duration time. In conclusion, although MSADC is not superior to entropy degree centrality in contact times, it has best performance in neighbor number, and precedes degree centrality in the new factors.

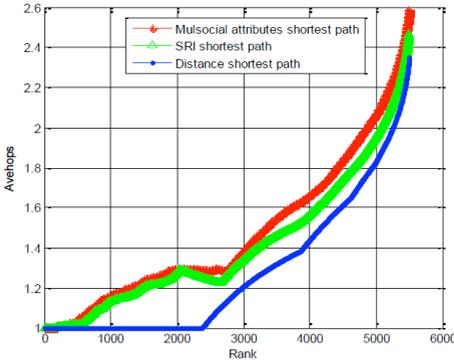


**Fig. 6.** Average adjacent nodes in MSADC ranking top N.

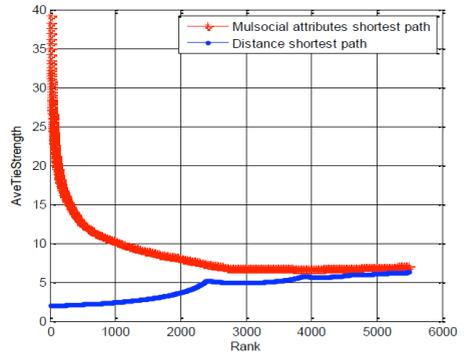


**Fig. 7.** Average duration time in MSADC ranking top N.

#### 4.4 Experimental results of multi-social attributes degree centrality (MSADC)



**Fig. 8.** Average hops in shortest distance ranking top N.



**Fig. 9.** Average weight in shortest distance ranking top N.

Closeness centrality and betweenness centrality rely on shortest paths among nodes to identify influential nodes in the network. In order to evaluate the effectiveness of the proposed centrality metrics in MSNs, we first evaluate the effectiveness of the proposed MSASP by MIT data set. We compare MSASP with SRI shortest path [16] and distance shortest path [17]. Figure 8 shows the the curves of average hops of the top-N nodes which are ranked shortest distance by MSASP, SRI shortest path, and distance shortest path. Because the distance between nodes is shorter, the path has less hops and higher weight. Figure 8 clearly shows that the average hops in MSASP are very close to SRI shortest path, and they are both higher than distance shortest path. Due to measure standards are different between tie strength and SRI, we cannot contrast their values. So there only exist MSASP and distance shortest path to compare the average weight between nodes ranking top N in figure 9. We obviously learn about MSASP is higher than distance shortest path. Above all, MSASP can better measure the shortest distance between nodes.

### 5 Conclusion

In mobile social networks, the sum of various human relations form social attributes, so the human mobility bring fierce impact. How to find the influential node in MSNs is the focus for many researchers. Therefore, this paper proposes a centrality measurement method based on multi-social attributes weighted for MSNs. Firstly, using the temporal evolution graph model, we quantify human social relations and mobility model as weight for the links and the experiment prove that mesurement of tightness proposed in this paper can consider more comprehensive factors and get more smooth curse compared with SRI. Then MSASP is come up with, and the experiment shows it can better measure the shortest distance between nodes. What's more, we redefine degree of centrality and its superiority is evaluated in the real data set.

As future work, we are planning to analyze the correlation between multi-social attributes degree centrality and multi-social attributes closeness and betweenness centralities. We will also discuss other data sets for the applicability.

### References

1. D. M. Boyd and N. B. Ellison, *Comput.-Mediated Commun.*, **13**, 210, ( 2007).
2. X Hu, T H S Chu, et al. *IEEE Communi. Surveys & Tutorials*, **17**, 1557(2015).

3. X Wang, H Wang, et al. *IEEE Inter. Conf. on Sen., Communi., and NET.* 1(2017).
4. W Stanley, F Katherine. New York: *Cambridge University Press.* 17(1994)
5. Q Xingqin, P Eddie, et al. *Information Sciences.* **194**, 240(2012).
6. Y Yustiawan, W Maharani, A Gozali. *Pro. Com.* **59**, 419(2015), .
7. T. Opsahl, F. Agneessens, J. Skvoretz, *Soc. Networks* **32**, 245 (2010).
8. E M Daly, M Haahr. *IEEE Trans. Mobile Computing* **8**, 606 (2009)
9. H Kim, R Anderson. *Phys. Rev. E*, **85**, 026107(2012).
10. V Nicosia, J Tang, C Mascolo, et al. *Temp. Net. Springer Berlin Heidelberg*, **15**(2013).
11. C Magnien, F Tarissan, 1200 (2015).
12. E W Dijkstra, *Numerische Mathematik*, **1**, 269 (1959).
13. M. E. J. NEWMAN, *Phys. Rev. E*, **64**, 016132. (2001)
14. A Barrat, M Barthélemy, et al. *Proce. National Academy. Sciences*, **101**, 37472004(2004).
15. S Basagni, M Conti, S Giordano, et al. *Wiley-IEEE Press*, 360-418(2013).
16. Gao Zhenxiang, Shi Yan, Chen Shanzhi, *Inte. Jou. Mod. Phys. C*, **26**, 1550107(2015).
17. L. C. Freeman, *Soc. Networks* 1, 215 (1979).
18. N. Eagle, A. S. Pentland and D. Lazer, *Proc. Nat. Acad. Sci.* **106**, 15274 (2009).