

Predicting temporal centrality in Opportunistic Mobile Social Networks based on social behavior of people

Huan Zhou¹ · Linping Tong¹ · Shouzhi Xu¹ · Chungming Huang² · Jialu Fan³

Received: 31 January 2016 / Accepted: 29 July 2016 / Published online: 20 September 2016
© Springer-Verlag London 2016

Abstract Predicting the centrality of nodes is a significant problem for different applications in Opportunistic Mobile Social Networks (OMSNs). However, when calculating such metrics, current studies focused on analyzing static networks that do not change over time or using aggregated contact information over a period of time. Furthermore, the centrality measured in the past is not verified whether it is useful as a predictor for the future. In this paper, in order to capture the dynamic behavior of people, we focus on predicting nodes' future centrality (importance) from the temporal perspective using real mobility traces in OMSNs. Three important centrality metrics, namely betweenness, closeness, and degree centrality, are considered. Through real trace-driven simulations, we find that nodes' future centrality is highly predictable due to natural social behavior of people. Then, based on the observations in the

simulation, we design several reasonable prediction methods to predict nodes' future temporal centrality. Finally, extensive real trace-driven simulations are conducted to evaluate the performance of our proposed methods. The results show that the Recent Weighted Average Method performs best in the MIT Reality trace, and the recent Uniform Average Method performs best in the Infocom 06 trace. Furthermore, we also evaluate the impact of parameters m and w on the performance of the proposed methods and find proper values of different parameters for each proposed method at the same time.

Keywords Opportunistic Mobile Social Networks · Centrality · Real mobility trace · Prediction method

1 Introduction

With the rapid development of wireless portable devices (e.g., smartphones, iPad, PDAs) with Bluetooth or Wi-Fi, Opportunistic Mobile Social Networks (OMSNs) begin to emerge [1–5]. OMSNs combine opportunistic mobile networks and mobile social networks together. Previous studies have shown that the performance of such networks depends highly on the user's social behavior as opportunistic mobile networks and mobile social networks share many common characteristics [6–8]. Their common features motivate an increasing research interests in OMSNs, especially using the social network analysis technology to help the design of routing protocols [9–12].

In order to use the social network analysis technology to design routing protocols in OMSNs, a significant problem is to measure the centrality (importance) of nodes in networks. Depending on the application, previous studies have proposed diverse metrics to measure the relative

✉ Shouzhi Xu
xsx@ctgu.edu.cn

Huan Zhou
zhouhuan117@gmail.com

Liping Tong
tlp496531232@163.com

Chungming Huang
huangcm@locust.csie.ncku.edu.tw

Jialu Fan
fanjialu@gmail.com

¹ College of Computer and Information Technology, China Three Gorges University, Yichang, China

² Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan

³ The State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, China

importance of nodes in networks such as betweenness centrality, closeness centrality, and degree centrality [13–16]. Betweenness centrality measures the extent to which a node lies on the shortest paths linking other nodes in the network, which has been applied to the study of identifying bottlenecks in traffic networks. Closeness centrality measures the distance a certain node to all other reachable nodes in the network, which has been applied to the study of information spreading. Finally, degree centrality is measured as the number of direct links with other nodes of a certain node, which can be applied to attack networks.

However, when calculating such centrality metrics, the current studies focused on analyzing static networks that do not change over time or using aggregated contact information over a period of time. Actually, nodes in OMSNs are inherently dynamic, which is driven by natural social behavior of people. Therefore, it is not prudent to assume stationary behavior of people in the design of practical applications. In particular, many researchers have also observed that nodes in OMSNs have a regular pattern [17–20]. For example, a student will go to school with his neighbors at morning every day and have classes with his classmates in the classroom, which brings a regular pattern of physical contacts, etc., which in turn provides the periodicity seen in the underlying communication processes. Here, we make a simple assumption: Since a node's schedule is regular, if it is an important node in the network in the past, then it is highly possible that its importance in the future will be correlated with the importance in the past.

Based on the assumption, we therefore focus on predicting nodes' future centrality from the temporal perspective under three important centrality metrics, namely betweenness, closeness, and degree centrality, using different real mobility traces in OMSNs. Our contributions in this paper are threefold:

1. Through real trace-driven simulations, we find that nodes' temporal centrality values are highly correlated with their past recent centrality values, and also have periodical behavior at 24 h difference.
2. Based on the observations, several intuitive reasonable prediction methods, e.g., Last Method, Recent Uniform Average Method, Recent Weighted Average Method, Periodical Average Method and Periodical Weighted Average Method, were designed to predict nodes' future temporal centrality.
3. The performance of the proposed prediction methods is evaluated using different real mobility traces, and the simulation results show that the best-performing prediction methods are more accurate on average than just using the Last Method. Moreover, we also evaluate

the impact of different parameters on the proposed prediction methods and suggest proper values of different parameters for the proposed prediction methods.

The remainder of this paper is organized as follows. We present the related work in Sect. 2 and the network model in Sect. 3. Then, we introduce the preliminaries in Sect. 4. Section 5 analyzes the correlation between the past and future temporal centrality value and proposes several methods to predict the future temporal centrality in OMSNs. Extensive simulations are conducted to evaluate the performance of the proposed methods in Sect. 6. At last, we conclude the paper in Sect. 7.

2 Related work

Centrality has been well investigated in OMSNs, which is widely applied to design social-based forwarding schemes in OMSNs. The basic idea is to use node centrality as the forwarding metrics, and the forwarding strategy is to forward data to nodes which have centrality value than the current node. SimBet [21] uses the egocentric betweenness metric and social similarity-based routing protocol to increase the performance of data delivery. BUBBLE Rap [22] considers node egocentric betweenness based on social communities to increase the data forwarding performance. Gao et al. [14] proposed the cumulative contact probability (CCP) as the centrality metrics to select relay nodes for multicasting in OMSNs. Since central nodes consume energy more quickly than other nodes, Socievole et al. [23] proposed an energy-aware centrality metric for information forwarding, which uses the node energy level to avoid transmissions in nodes more central. Based on degree centrality and betweenness centrality, Zhu et al. [24] proposed a set of social-based throwbox placement algorithms to efficiently deploy throwboxes in the network. However, the above studies use static networks that do not change over time or aggregated contact information over a period of time to calculate the centrality value. Furthermore, they do not verify whether the centrality measured in the past is useful as a predictor for the future.

Currently, some studies have attempted to capture the dynamic behavior of people [25–27]. Chaintreau et al. [25] found that the pairwise inter-contact time can be well approximated by a power-law distribution. Karagiannis et al. [26] found that the pairwise inter-contact time follows a power-law distribution, with an exponential cutoff. Passarella et al. [27] found that the aggregate inter-contact times distribution is not representative, in general, of the individual pairs distributions. These results above have

been used to model potential future contact patterns but do not provide much insight into the problem of predicting future network structure. To capture the dynamic behavior of people more accurately, some studies have tried to solve the problem from the temporal perspective [28–32]. Tang et al. [28] proposed three temporal centrality metrics based on temporal paths in order to measure the importance of a node in a dynamic network. Kim et al. [29] proposed several methods to predict nodes’ importance in the future. Huang et al. [31] proposed a kernel regression-based estimation framework for link pattern prediction in OMSNs. In order to investigate the evolution of dynamic social networks, Wei et al. [32] proposed two classes of dynamic metrics in terms of persistency and emergence, and used three different temporal aggregation models to implement them. Our work is similar to the work in [29], but we ignore the lagged time, and only predict the centrality value of a single time window in the network model, which makes the problem more clear and the prediction results more reasonable.

3 Network model

In this section, we introduce the network model used in this paper. We assume that the time during which a network is observed is finite, from t_{start} to t_{end} ; without loss of generality, we set $t_{\text{start}} = 0$ and $t_{\text{end}} = T$. A dynamic network contact graph $G_{0,T} = (V, E_{0,T})$ on a time interval $[0, T]$ consists of a set of vertices V and a set of temporal edges $E_{0,T}$, where stochastic contact process between a node pair $u, v \in V$ on a time interval $[t_s, t_e]$ ($0 \leq t_s \leq t_e \leq T$) is modeled as an temporal edge $e_{t_s,t_e}^{uv} \in E_{0,T}$. The characteristics of a temporal edge are mainly determined by the properties of the inter-contact time and contact duration among mobile nodes.

Most characterizations of dynamic networks discretize time by converting temporal information into a sequence of network “snapshots” to apply techniques derived from graph theory to the analysis of networks. As shown in Fig. 1, the time interval $[0, T]$ is divided into fixed discrete time windows $\{1, 2, \dots, n\}$. We use $w = \frac{T}{n}$ to denote the size of each time window, expressed in some time unites (e.g., minutes or hours). In other words, a dynamic network can be represented as a series of static graphs at each time,

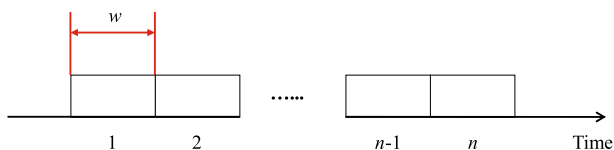


Fig. 1 A Series of “snapshots”

G_1, G_2, \dots, G_n . The notation G_t ($1 \leq t \leq n$) represents the aggregate graph which consists of a set of vertices V and a set of edges E_t where an edge $e_{t_s,t_e}^{uv} \in E_t$ exists only if a temporal edge $e_{t_s,t_e}^{uv} \in E_{0,T}$ exists between vertices u and v on a time interval such that $t_e \leq wt$ and $t_s > w(t - 1)$. In other words, G_t is the t th temporal snapshot of the dynamic contact graph $G_{0,T}$ during the t th time window.

Figure 2 shows the comparison between the aggregated static graph and the time-varying dynamic graph, in which $t_{\text{start}} = 0, t_{\text{end}} = 3$ and $w = 1$. Unlike the aggregated view of the static graph shown in Fig. 2a, a series of static contact graphs $G_1, G_2,$ and G_3 in Fig. 2b represent temporal edge relationships between nodes $A, B, C,$ and D . For example, if we look at the aggregated static graph in Fig. 2a, node A has a contact edge with node C ; however, if we look at the temporal static graph in G_1 and G_2 , node A does not have contact edge with node C . Therefore, the contact edge in the temporal static graph is obviously different from that in the aggregated static graph. In the next section, we will focus on investigating the centrality based on the temporal static graph.

4 Preliminaries

In this section, we first define definition and notation related to the centrality metrics and then introduce the centrality prediction problem which will be solved in the rest of the paper.

4.1 Network centrality measures

Centrality refers to a group of metrics that aim to quantify the “importance” or “influence” of a particular node (or group) within a network. There are several common

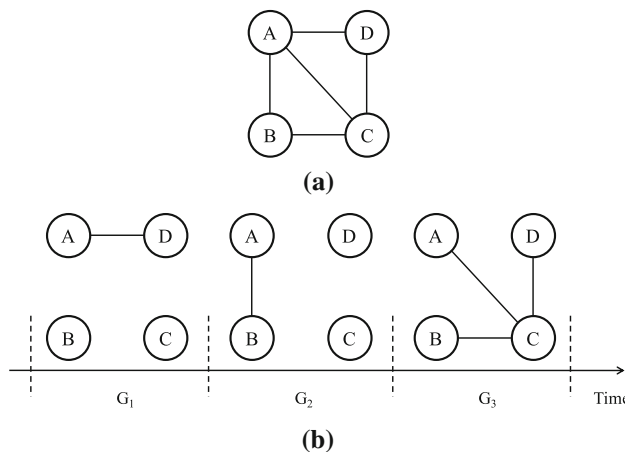


Fig. 2 Comparison between the aggregated static graph and the time-varying dynamic graph. a Aggregated static graph; b time-varying dynamic graph

methods to measure “centrality.” In this paper, we only introduce three of them: betweenness, closeness, and degree centrality. Formally, we use the standard definition of the betweenness, closeness, and degree centrality, and the centrality value of a node i can be expressed as follows [13].

4.1.1 Betweenness centrality

Betweenness centrality measures the extent to which a node lies on the shortest paths linking other nodes in the network, which is calculated as the proportional number of shortest paths between all node pairs in the network, that pass through a certain node. Betweenness centrality of a certain node i can be expressed as:

$$\text{Betweenness}(i) = \sum_{u \neq i, v \neq i, i \in V}^r \frac{\delta_{u,v}(i)}{\delta_{u,v}} \quad (1)$$

where $\delta_{u,v}$ is the total number of shortest paths starting from the source node u and the destination node v , and $\delta_{u,v}(i)$ are the number of shortest paths starting from the source node u and the destination node v which actually pass through node i .

4.1.2 Closeness centrality

Closeness centrality measures the distance a certain node to all other reachable nodes in the network, which is calculated as the average shortest path length between a certain node and all other reachable nodes. Closeness centrality can be regarded as a measure of how long it will take message to spread from a given node to other nodes in the network. Closeness centrality of a certain node i can be expressed as:

$$\text{Closeness}(i) = \frac{1}{|V| - 1} \sum_{j \neq i, j \in V}^r \Delta_{i,j}(i) \quad (2)$$

where $\Delta_{i,j}$ is the number of hops in the shortest path from node i to node j and V is the set of nodes in the network.

4.1.3 Degree centrality

Degree centrality is measured as the number of direct links with other nodes of a certain node. A node with high degree centrality has numerous contacts with other nodes in the network. Degree centrality of a certain node i can be calculated as:

$$\text{Degree}(i) = \frac{\sum_{j \neq i, j \in V}^r e(i,j)}{|V| - 1} \quad (3)$$

where $e(i, j) = 1$ if a direct link exists between node i and j , and V is the set of nodes in the network.

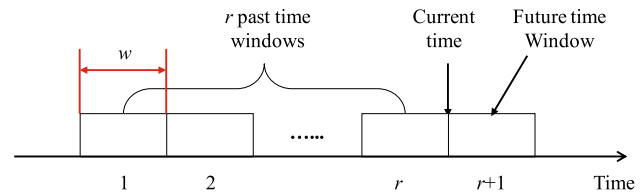


Fig. 3 Illustration of the past and future time windows

4.2 Centrality prediction problem

As shown in Fig. 3, we generalize the problem for centrality prediction in this paper as follows: Given a dynamic network $G_{1,r}$ observed during r past time intervals, we want to predict the average network centrality values of the nodes in the network in the $r + 1$ time intervals. Therefore, the purpose of this paper is to propose several prediction methods to minimize the prediction error, and evaluate the impact of different parameters on the performance of the proposed prediction methods. In order to evaluate the effectiveness of the prediction methods, we use $\text{Error}(G_t)$ to denote the average error between the guessed centrality values and the true centrality values, which can be expressed as:

$$\text{Error}(G_t) = \frac{\sum_{i \in V} |C_t(i) - \hat{C}_t(i)|}{|V|} \quad (4)$$

where $C_t(i)$ is node i 's centrality value such as betweenness(i), closeness(i), or degree(i) in G_t , and $\hat{C}_t(i)$ to denote the node i 's predicted centrality value in G_t .

5 Centrality prediction

In this section, we first use two real mobility traces, *Infocom 06* [33] and *MIT Reality* [34] to test whether the centrality can be predicted. Then, based on the findings, several methods were introduced to predict the future temporal centrality.

5.1 Analysis of correlation between past and future centrality

Since people in reality always have regularity, we hypothesize that the past centrality has high correlation with the future centrality. To test this hypothesis, we use two real mobility traces, *Infocom 06* [33] and *MIT Reality* [34] collected from real environments. Users in these two traces are all carrying Bluetooth-enabled portable devices, which record contacts by periodically detecting their peers nearby. The traces cover various types of corporate environments and have various experiment periods. The details of the traces are summarized in Table 1. We use

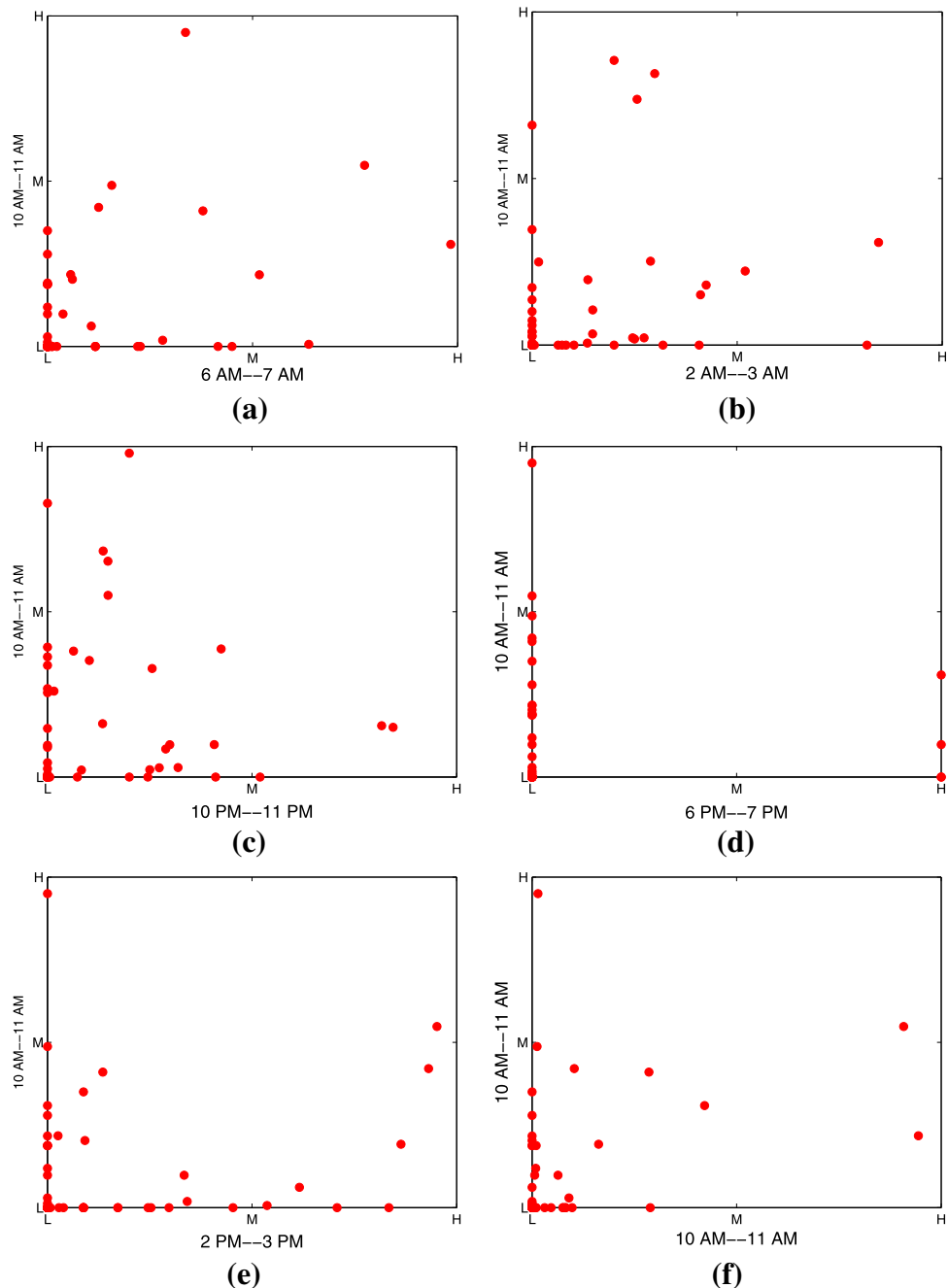
Table 1 Trace statistics

Trace	MIT Reality	Infocom 06
Device	Smartphones	iMote
Network type	Bluetooth	Bluetooth
Duration (days)	246	4
Granularity (s)	300	120
No. of internal contacts	114,046	182,951
No. of devices	97	78
Contact frequency/pair/day	0.024	6.7

part of traces (a typical weekday in October of MIT Reality trace, the second day of Infocom 06 trace) to test the hypothesis.

Figure 4 shows each node’s betweenness centrality value compared to its values in past windows, in the MIT Reality trace. We find that there is a high correlation (0.4733) between a node’s temporal betweenness centrality value with its value 4 h ago; second, increasing the time difference decreases the correlation (e.g., 0.3407 at –8 h time difference and 0.2096 at –12 h time difference); and

Fig. 4 Scatter plots depicting betweenness centrality correlation between a fixed window (y-axis) and an increasingly distant window from the past (x-axis) every 4 h in the MIT trace. The axis labels represent the low (L), medium (M), and high (H) betweenness centrality values. **a** Time difference: –4 h, average correlation: 0.4733, **b** time difference: –8 h, average correlation: 0.3407, **c** time difference: –12 h, average correlation: 0.2096, **d** time difference: –16 h, average correlation: 0.0625, **e** time difference: –20 h, average correlation: 0.2565, **f** time difference: –24 h, average correlation: 0.4155



third, at -24 h time difference the correlation rises again (0.432), which indicates possible periodic behavior.

Figure 5 shows each node's betweenness centrality value compared to its values in past windows, in the Infocom 06 trace. We find that similar to the results in the MIT Realty trace, recent past betweenness centrality values are highly correlated compared with more distant values (e.g., 0.5804 at -4 h time difference and 0.2451 at -8 h time difference). However, different from the results in MIT Realty trace, the pattern of repeated peaks

with -24 h time difference seems rather weak in Infocom 06 trace. The main reason is that people attending the IEEE Infocom 2006 conference are more likely to seek out new colleagues to talk to at the breaks between sessions, rather than socializing with the same people, but students in the MIT campus are more likely to meet the same people when they are taking classes or walking in the campus.

Figure 6 shows each node's closeness centrality value compared to its values in past windows, in the MIT Realty

Fig. 5 Scatter plots depicting betweenness centrality correlation between a fixed window (y-axis) and an increasingly distant window from the past (x-axis) every 4 h in the Infocom 06 trace. The axis labels represent the low (L), medium (M), and high (H) betweenness centrality values. **a** Time difference: -4 h, average correlation: 0.5804, **b** time difference: -8 h, average correlation: 0.2451, **c** time difference: -12 h, average correlation: 0.0731, **d** time difference: -16 h, average correlation: 0.0079, **e** time difference: -20 h, average correlation: -0.0155 , **f** time difference: -24 h, average correlation: 0.0254

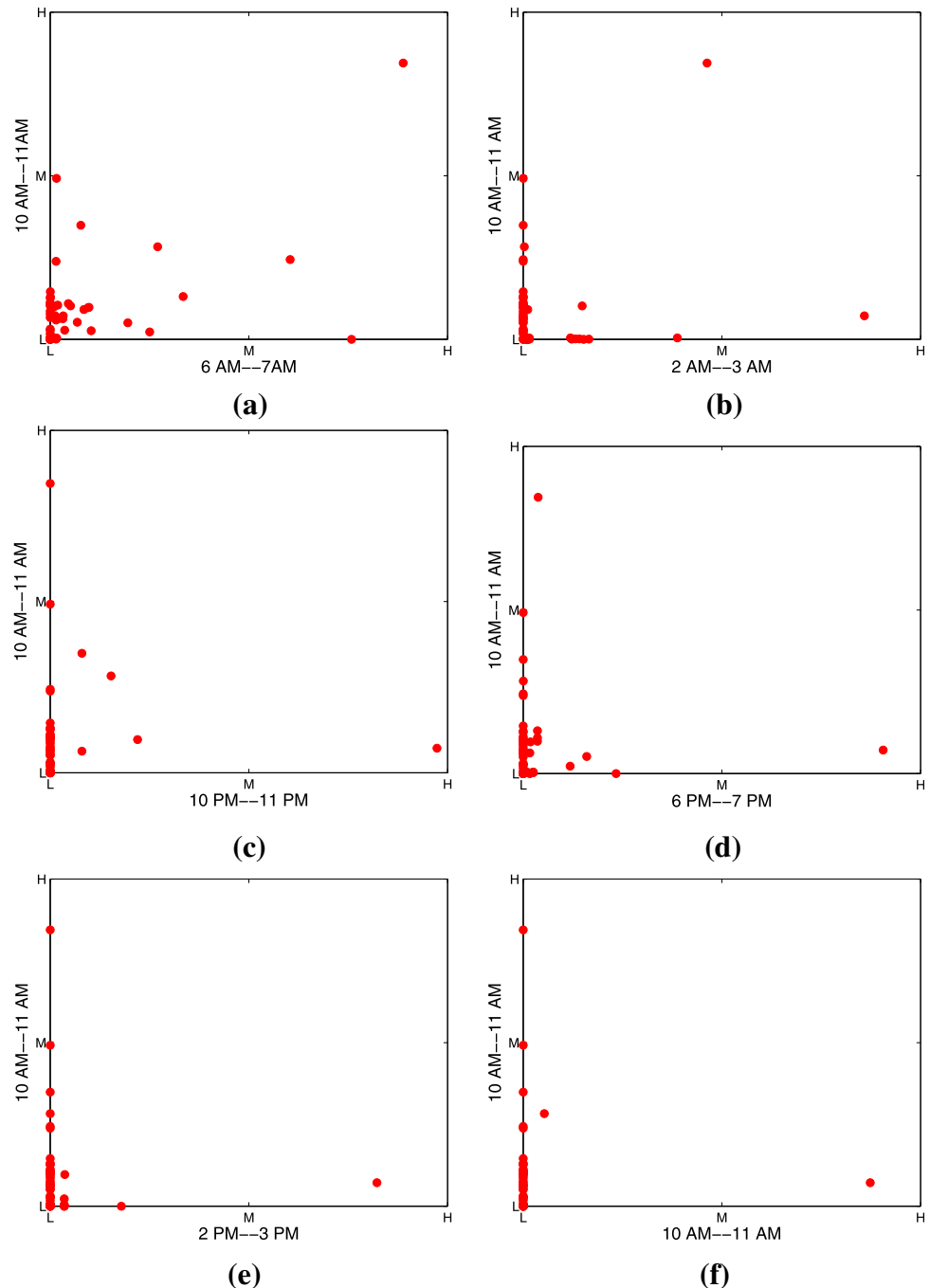
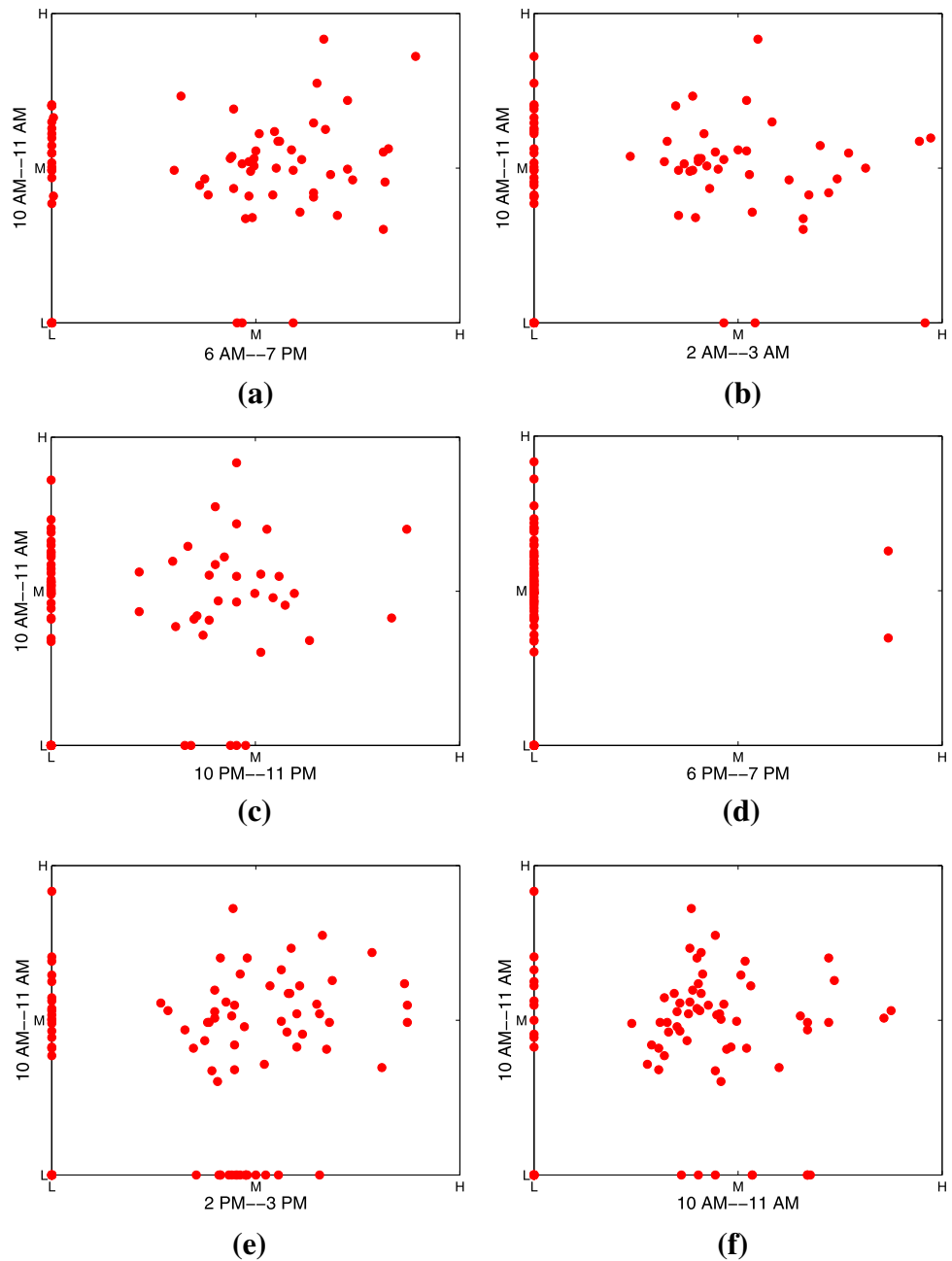


Fig. 6 Scatter plots depicting closeness centrality correlation between a fixed window (y-axis) and an increasingly distant window from the past (x-axis) every 4 h in the MIT Reality trace. The axis labels represent the low (L), medium (M), and high (H) closeness centrality values. **a** Time difference: -4 h, average correlation: 0.5207, **b** time difference: -8 h, average correlation: 0.3534, **c** time difference: -12 h, average correlation: 0.2741, **d** time difference: -16 h, average correlation: 0.00726, **e** time difference: -20 h, average correlation: 0.2433, **f** time difference: -24 h, average correlation: 0.3778



trace. Similar to the results in Fig. 4, there is also a high correlation (0.5207) between a node’s temporal closeness centrality value with its value 4 h ago; second, increasing the time difference also decreases the correlation (e.g., 0.3534 at -8 h time difference and 0.2741 at -12 h time difference); and third, at -24 h time difference the correlation rises again (0.3778). Figure 7 shows each node’s degree centrality value compared to its values in past windows, in the Infocom 06 trace. Similar to the results in Fig. 5, recent past degree centrality values are also highly correlated compared with more distant values (e.g., 0.4858 at -4 h time difference and 0.3252 at -8 h time

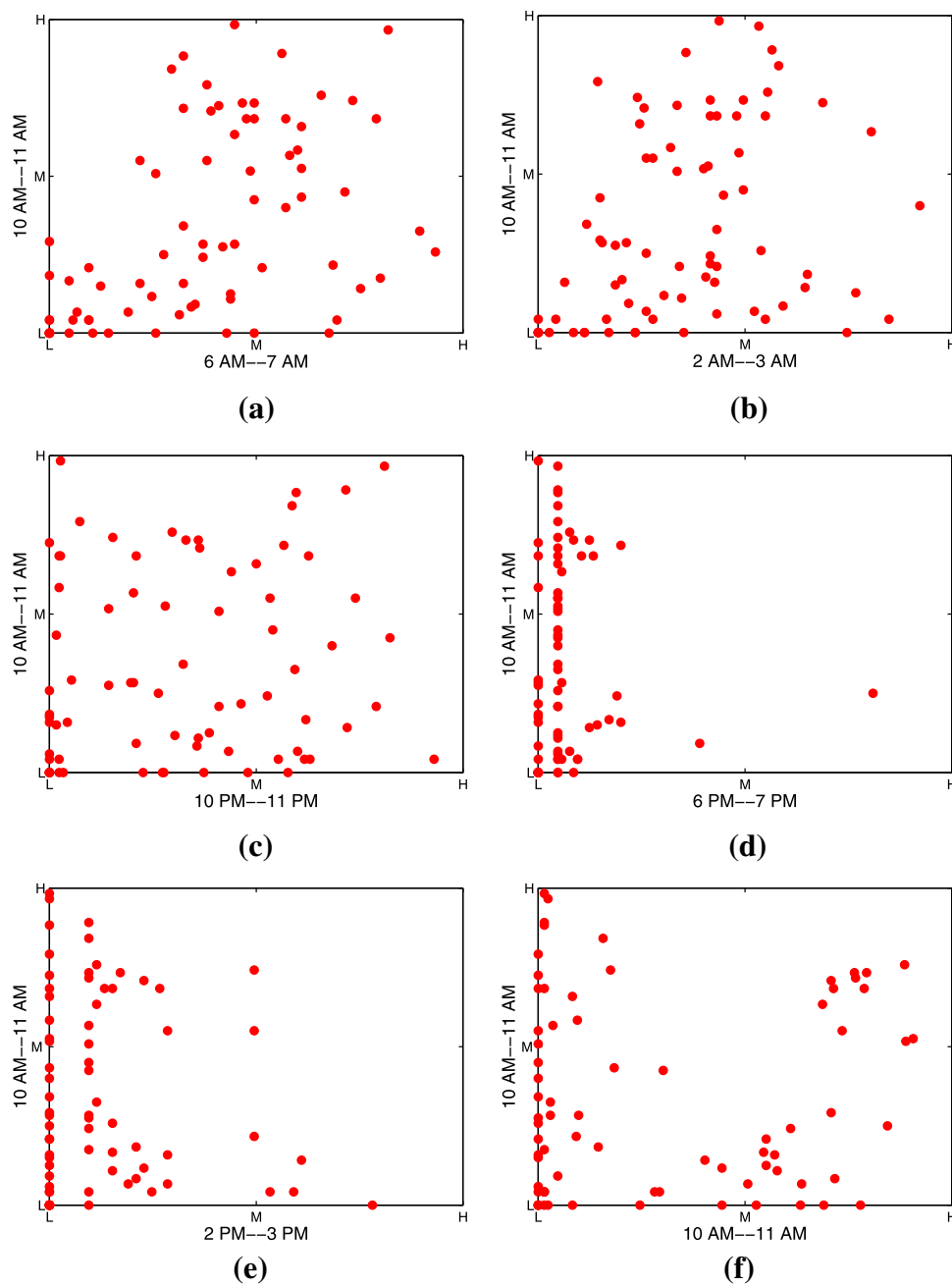
difference), and the pattern of repeated peaks with -24 h time difference seems also very weak in Infocom 06 trace.

Based on the results reported above, we have two key observations:

1. Recent past centrality values are highly correlated compared with more distant values in 1 day.
2. A node’s temporal centrality value with its value at 24 h difference are highly correlated, which indicates possible periodic behavior.

In the next part, we will present several prediction methods based on these observations.

Fig. 7 Scatter plots depicting degree centrality correlation between a fixed window (y-axis) and an increasingly distant window from the past (x-axis) every 4 h in the Infocom 06 trace. The axis labels represent the low (L), medium (M), and high (H) degree centrality values. **a** Time difference: -4 h, average correlation: 0.4858, **b** time difference: -8 h, average correlation: 0.3252, **c** time difference: -12 h, average correlation: 0.1511, **d** time difference: -16 h, average correlation: 0.0009, **e** time difference: -20 h, average correlation: -0.0699, **f** time difference: -24 h, average correlation: 0.056



5.2 Prediction methods

Based on the analysis above, in this part, we introduce several methods to predict the future temporal centrality.

Last Method As the first candidate, we just use the node's temporal centrality value in the last temporal network (G_r) at time window r . In other words, for $i \in V$, we use the temporal centrality $C_r(i)$ in G_r as the future temporal centrality $C_{fu}(i)$ in G_{r+1} .

Recent Uniform Average Method In order to improve the accuracy of the prediction, we can use the node's m previous centrality values instead of one last previous

centrality value. A reasonable idea is to use the node's uniform average centrality value between $G_{r-m+1}, \dots, G_{r-1}, G_r$, where $0 < m \leq r$ as the node's future temporal centrality value. Formally, the future temporal centrality $C_{fu}(i)$ can be expressed as:

$$C_{fu}(i) = \frac{1}{m} \sum_{k=r-m+1}^r C_k(i) \quad (5)$$

We want to find the best m given the cost of computation and the accuracy of prediction, and will suggest values based on different real traces.

Recent Weighted Average Method In order to consider the relative importance of the recent temporal networks, we can use the recent weighted average centrality value instead of the uniform average centrality value. Therefore, the future temporal centrality $C_{fu}(i)$ using the recent weighted average centrality method can be expressed as:

$$C_{fu}(i) = \sum_{k=r-m+1}^r C_k(i)W_d/W_{total} \tag{6}$$

where $0 < m \leq r$; $d = r + 1 - k$ is the time difference between the past k th time window and the future time window $r + 1$; $W_{total} = \sum_{k=r-m+1}^r W_d$. In fact, the recent uniform average centrality is a special case of the recent weighted average centrality when $W_d = 1/m$. Here, we consider the linear ($W_d = \frac{1}{d}$) weight assignments depending on the time difference d between G_k and G_{r+1} .

Periodical Average Method According to the analysis above, we find that activities of people are repeated periodically. Hence, an intuitive method is to use these periodical patterns to improve the accuracy of the prediction. For contact network of people, reasonable periods are a day or week. Given the period p of a day or a week, we consider using the node’s periodical average centrality value between $G_{r-m+1}, \dots, G_{r-1}, G_r$ where $0 < m \leq r$ as the node’s future temporal centrality value. Hence, we first have to find periodical time windows of the $(r + 1)$ th time window. We define $a = \min_{k=r-m+1}^r (r + 1 - k)w \bmod p$ as the time window which is the closest to the periodical time window of the $(r + 1)$ th time window. Then, we define $f(k)$ as:

$$f(k) = \begin{cases} 1, & kw \bmod p \equiv a, \\ 0, & kw \bmod p \not\equiv a. \end{cases} \tag{7}$$

Then, the future temporal centrality $C_{fu}(i)$ in G_{r+1} can be expressed as:

$$C_{fu}(i) = \frac{\sum_{k=r-m+1}^r f(k)C_k(i)}{\sum_{k=r-m+1}^r f(k)} \tag{8}$$

Periodical Weighted Average Method In order to consider the relative importance of the recent temporal networks, similar to the method above, we can use the recent periodical weighted average centrality value instead of the periodical uniform average centrality value. Therefore, based on Eqs. 7 and 8, the future temporal centrality $C_{fu}(i)$ in G_{r+1} using the recent periodical weighted average centrality method can be expressed as:

$$C_{fu}(i) = \frac{\sum_{k=r-m+1}^r f(k)W_dC_k(i)}{\sum_{k=r-m+1}^r W_{total}} \tag{9}$$

where $d = (r + 1 - k)/p$ is the periodical time difference between the past k th time window and the future time

window $r + 1$; $W_{total} = \sum_{k=r-m+1}^r f(k)W_d$. Similar to the method above, we consider the linear ($W_d = \frac{1}{d}$) weight assignments depending on the periodical time difference d between G_k and G_{r+1} .

6 Performance evaluation

In this section, we aim to evaluate the performance of the proposed methods above in the Infocom 06 and MIT Reality traces, and the impact of several parameters (e.g., m and w) on the performance of the proposed methods above. We use part of MIT Reality traces (the first 10 days of October), and all traces in the Infocom 06 trace to evaluate the performance of our proposed methods. For each prediction method, $Error(G_t)$ is used to evaluate the prediction accuracy of the proposed methods. Here, we use Last to denote the Last Method, RUA to denote the Recent Uniform Average Method, RWA to denote the Recent Weighted Average Method, PUA to denote the periodical uniform average method, and PWA to denote the Periodical Weighted Average Method.

6.1 The impact of parameter m on the performance of the proposed methods

In this part, we focus on evaluating the impact of parameter m on the performance of the proposed methods, and find a proper value of m for each proposed method at the same time.

Figure 8 shows the temporal centrality prediction results of different prediction methods in the MIT Reality trace by varying m . We find that the RWA achieves the best results in closeness and degree, and is very close to the best results in betweenness. Moreover, RWA reaches the minimum $Error(G_t)$ more quickly than RUA, and is more stable than other methods. Therefore, we recommend to use RWA to predict the future temporal centrality in the MIT Reality trace. It is worth noticing that with the increasing of m , RWA achieves the minimum $Error(G_t)$ (below 0.0045 in closeness and 0.0113 in degree) when m is around 100 h in closeness and degree, and then the $Error(G_t)$ values will be stable after this time interval.

Although the performance of PWA is not as good as that of RWA, it outperforms the other prediction methods in closeness and degree. This is because the periodic patterns of nodes are clearly shown in the MIT Reality trace. Therefore, considering the performance of PWA, we recommend using PWA as an alternative. However, we would not recommend using Last Method because its relative accuracy is not enough, although its computation cost is relatively cheap.

Fig. 8 Temporal centrality prediction results by varying m in the MIT Reality traces when w is 1 h. **a** betweenness, **b** closeness, **c** degree

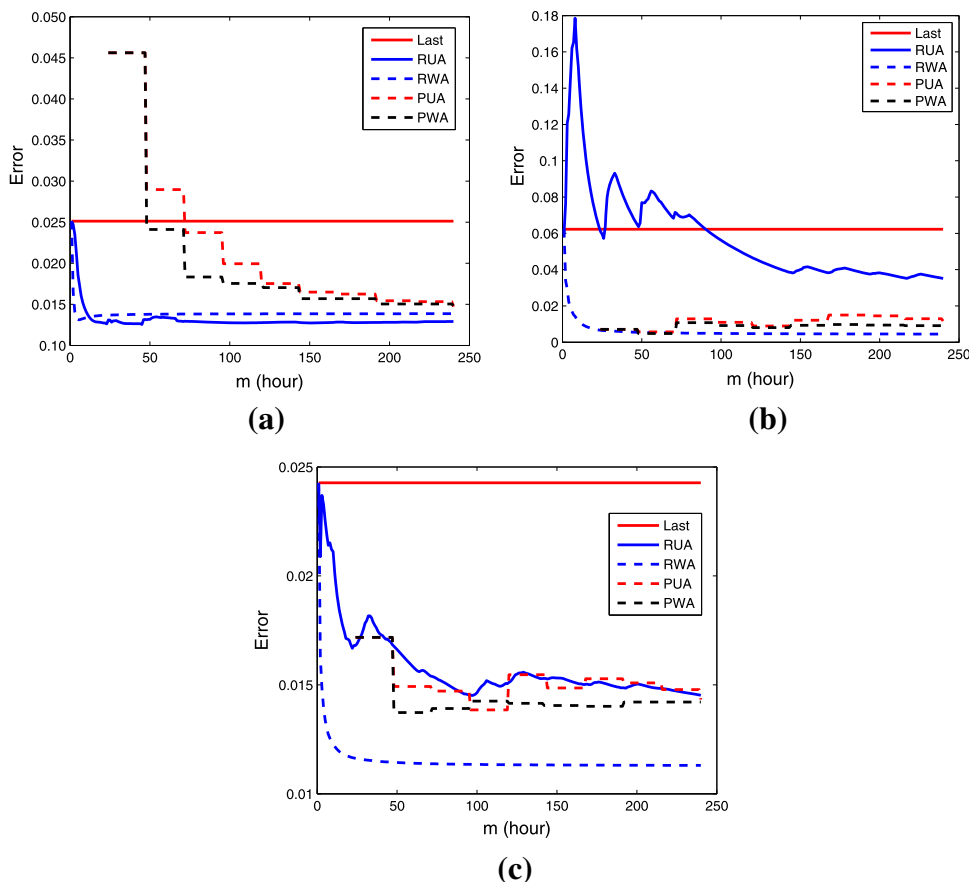


Figure 9 shows the temporal centrality prediction results of different prediction methods in the Infocom 06 trace by varying m . By contrast, we find that with the increasing of m , RWA and PWA are not as good as that in the MIT Reality trace, especially the PWA. Actually, we have already observed that there is no noticeable periodic patterns while the recent past centrality values are highly correlated in the Infocom 06 trace—Fig. 5 illustrates this. Therefore, we recommend to use RUA to predict the future temporal centrality value in the Infocom 06 trace. Furthermore, RUA achieves the minimum $Error(G_t)$ (0.1095 in betweenness, 0.1415 in closeness, and 0.0671 in degree) when m is 3 h, and then the $Error(G_t)$ values will increase after this time interval. This imply that the centrality value at a specific time in the Infocom 06 trace is highly related to the recent past centrality values in 3 h.

The performance of RWA is very close to that of RUA. Therefore, we recommend using RWA as an alternative. Similar to the results in the MIT Reality trace, we would not recommend using Last Method because its relative accuracy is not enough, although its computation cost is relatively cheap.

In summary, we recommend to use RWA and PWA to predict the future temporal centrality in the MIT Reality

trace, and use RUA and RWA to predict the future temporal centrality in the Infocom 06 trace. Moreover, the parameter m also has a significant impact on the performance of the proposed prediction methods. RWA in the MIT Reality trace will achieve the minimum $Error(G_t)$ and then be stable when m increases to a certain value, while RUA will achieve the minimum $Error(G_t)$ when m increases to a certain value, and then the $Error(G_t)$ values will increase after this time interval.

6.2 The impact of parameter w on the performance of the proposed methods

In this part, we focus on evaluating the impact of parameter m on the performance of the proposed methods, and find a proper value of w for each proposed method at the same time.

Figure 10 shows the temporal centrality prediction results of different prediction methods in the MIT Reality trace by varying w , and Fig. 11 shows the temporal centrality prediction results of different prediction methods in the Infocom 06 trace by varying w . We find that the error of all prediction methods (except RUA in the MIT Reality trace) increases with w increases, not only in the MIT

Fig. 9 Temporal centrality prediction results by varying m in the Infocom 06 traces when w is 1 h. **a** betweenness, **b** closeness, **c** degree

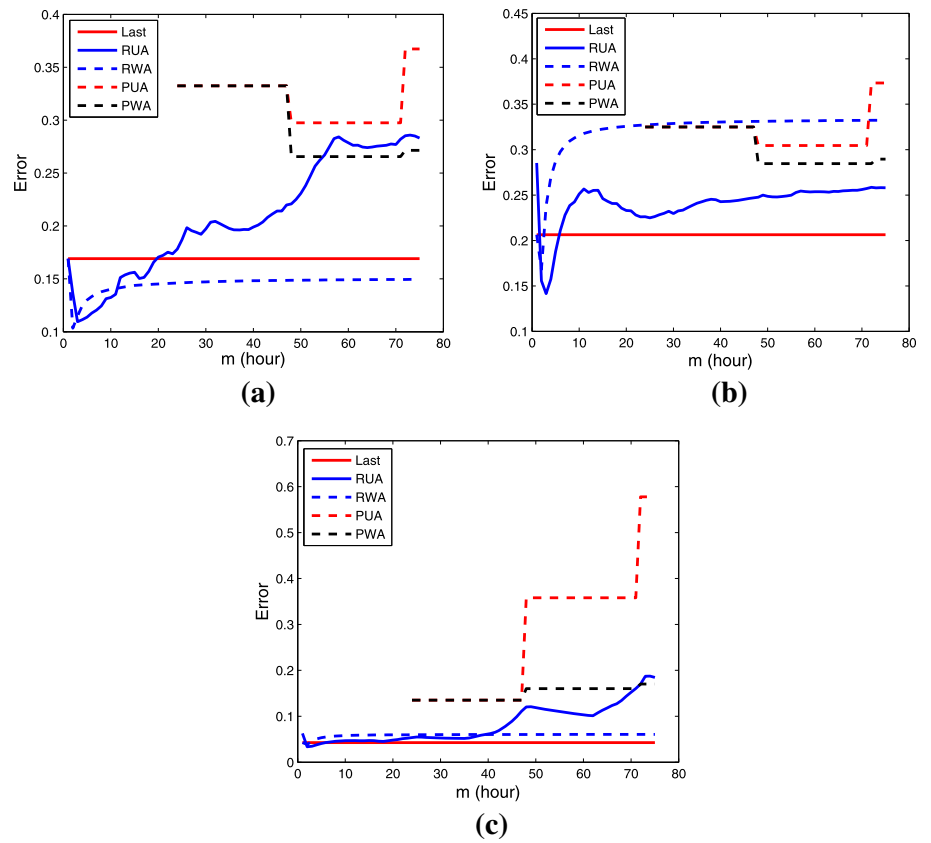


Fig. 10 Temporal centrality prediction results by varying w in the MIT Reality traces when m is 48 h. **a** betweenness, **b** closeness, **c** degree

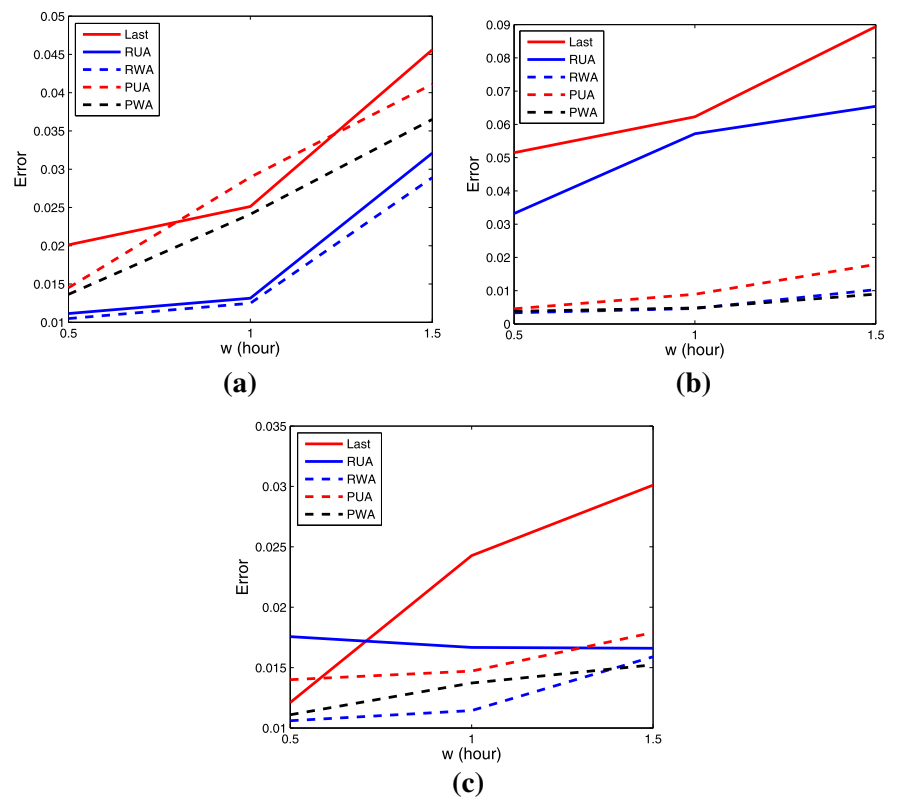
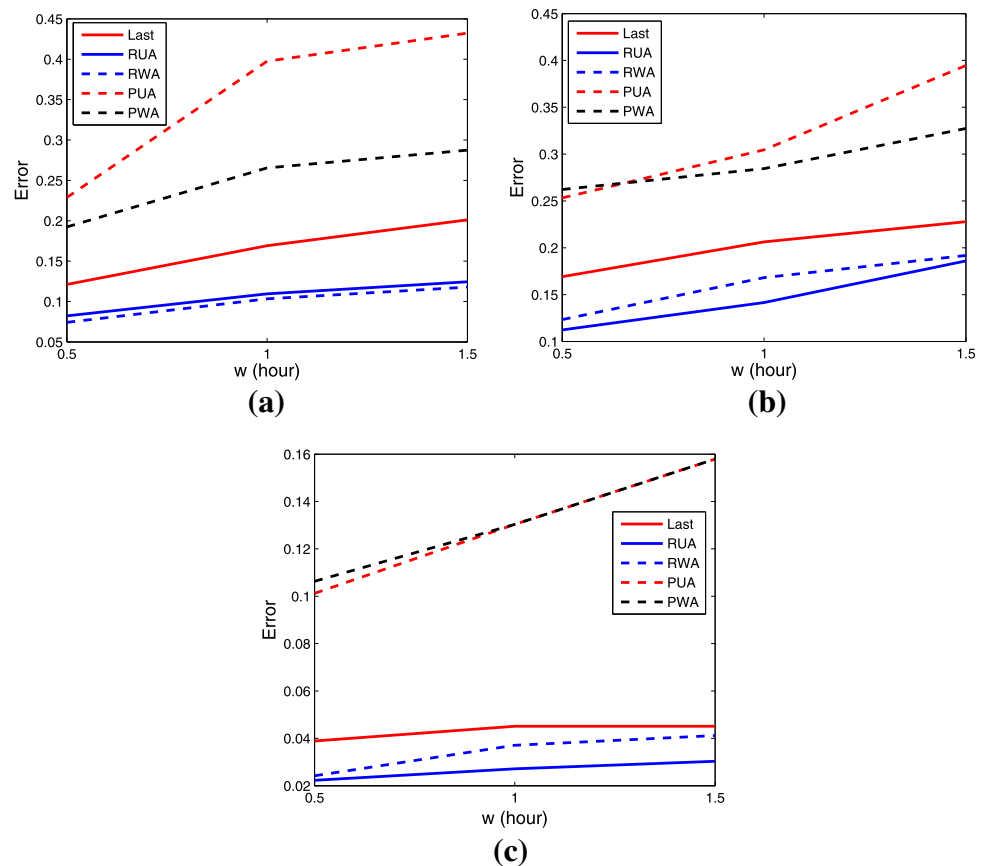


Fig. 11 Temporal centrality prediction results by varying w in the Infocom 06 traces when m is 48 h. **a** betweenness, **b** closeness, **c** degree



Reality trace, but also in the Infocom 06 trace. This is reasonable because a finer granularity may be desired to improve the prediction accuracy. However, w cannot be too small, because larger w means larger computation cost. Therefore, we should set the value of w according to the requirement of different applications. With the increase of w , RWA still performs best in the MIT Reality trace. Therefore, we still recommend to use RWA to predict the future temporal centrality in the MIT Reality trace. Similarly, RUA performs best in the Infocom 06 trace, and RWA is very close to the performance of RUA. Therefore, we recommend to use RUA or RWA to predict the future temporal centrality in the Infocom trace.

In summary, the parameter w has a significant impact on the performance of the proposed methods. Although the error of all prediction methods (except RUA in the MIT Reality trace) increases with the increase of w , w cannot be too small. Larger w means larger computation cost. We should set the value of w according to the requirement of different applications. Furthermore, with the change of w , we recommend to use RWA to predict the future temporal centrality in the MIT Reality trace, and RUA or RWA to predict the future temporal centrality in the Infocom trace.

7 Conclusions

In this paper, we have predicted nodes' future importance from the temporal perspective under three important metrics, namely betweenness, closeness, and degree centrality, using different real mobility traces in OMSNs. Through real trace-driven simulations, we find that nodes' centrality values are highly correlated with their past recent centrality values, and have periodical behavior at 24 h difference. Then, based on the observations in the simulation, we design several reasonable prediction methods to predict nodes' future temporal centrality. Extensive real trace-driven simulation results show that the Recent Weighted Average Method performs best in the MIT Reality trace, and the Recent Uniform Average Method performs best in the Infocom 06 trace. Furthermore, we also evaluate the impact of parameters m and w on the performance of the proposed prediction methods, and suggest proper values of parameters m and w for each proposed prediction method at the same time. In the future, we plan to analyze other real mobility traces to study their common characteristics, and use the predicted temporal centrality to design efficient routing protocol in OMSNs.

Acknowledgments This research was supported in part by NSFC under Grants 61602272, 61503147, and 41172298, the Open Research Project of State Key Laboratory of Synthetical Automation for Process Industries under Grant PAL-N201507, and Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering under Grant 2014KLA07.

References

- Fan J, Chen J, Du Y, Wang P, Sun Y (2011) Delque: a socially-aware delegation query scheme in delay tolerant networks. *IEEE Trans Veh Technol* 60(5):2181–2193
- Zhang D, Zhang D, Xiong H, Hsu C, Vasilakos A (2014) BASA: building mobile ad-hoc social networks on top of android. *IEEE Netw* 28(1):4–9
- Yu Q, Chen J, Fan Y, Shen X, Sun Y (2010) Multi-channel assignment in wireless sensor networks: a game theoretic approach. In: Proceedings of IEEE INFOCOM, pp 1–9
- He J, Cheng P, Chen J, Shi L, Lu R (2014) Time synchronization for random mobile sensor networks. *IEEE Trans Veh Technol* 63(8):3935–3946
- Zhou H, Chen J, Zheng H, Wu J (2016) Energy efficiency and contact opportunities tradeoff in opportunistic mobile networks. *IEEE Trans Veh Technol* 65(5):3723–3734
- Zhao D, Ma H, Tang S, Li X (2015) Coupon: a cooperative framework for building sensing maps in mobile opportunistic networks. *IEEE Trans Parallel Distrib Syst* 26(2):392–402
- Li F, Wu J (2009) MOPS: providing content-based service in disruption-tolerant networks. In: Proceedings of IEEE ICDCS
- Wang Z, Liao J, Cao Q, Qi H, Wang Z (2015) Friendbook: a semantic-based friend recommendation system for social networks. *IEEE Trans Mob Comput* 29(4):40–45
- Zhou H, Chen J, Zhao H, Gao W, Cheng P (2013) On exploiting contact patterns for data forwarding in duty-cycle opportunistic mobile networks. *IEEE Trans Veh Technol* 62(9):4629–4642
- Yuan Q, Cardei I, Wu J (2009) Predict and relay: an efficient routing in disruption-tolerant networks. In: Proceedings of ACM Mobihoc, pp 95–104
- Chen H, Lou W (2016) Contact expectation based routing for delay tolerant networks. *Ad Hoc Netw* 36:244–257
- Chen H, Lou W (2014) Gar: Group aware cooperative routing protocol for resource-constraint opportunistic networks. *Comput Commun* 48:20–29
- Scott J (1988) Social network analysis. *Sociology* 22(1):109–127
- Gao W, Li Q, Zhao B, Cao G (2009) Multicasting in delay tolerant networks: a social network perspective. In: Proceedings of ACM Mobihoc. ACM, pp 299–308
- Fan J, Chen J, Du Y, Gao W, Wu J, Sun Y (2013) Geo-community-based broadcasting for data dissemination in mobile social networks. *IEEE Trans Parallel Distrib Syst* 24(4):734–743
- Wang S, Huang L, Hsu C, Yang F (2016) Collaboration reputation for trustworthy web service selection in social networks. *J Comput Syst Sci* 82(1):130–143
- Hui P, Chaintreau A, Scott J, Gass R, Crowcroft J, Diot C (2005) Pocket switched networks and human mobility in conference environments. In: Proceedings of the ACM SIGCOMM workshop on Delay-tolerant networking. ACM, pp 244–251
- Zhou H, Chen J, Fan J, Du Y, Das SK (2013) ConSub: incentive-based content subscribing in selfish opportunistic mobile networks. *IEEE J Sel Areas Commun* 31(9):669–679
- Zhou H, Wu J, Zhao H, Tang S, Chen C, Chen J (2015) Incentive-driven and freshness-aware content dissemination in selfish opportunistic mobile networks. *IEEE Trans Parallel Distrib Syst* 26(9):2493–2505
- Zhao H, Zhou H, Yuan C, Huang Y, Chen J (2015) Social discovery: exploring the correlation among three-dimensional social relationships. *IEEE Trans Comput Soc Syst* 2(3):77–87
- Daly EM, Haahr M (2007) Social network analysis for routing in disconnected delay-tolerant manets. In: Proceedings of ACM MobiHoc, pp 32–40
- Hui P, Crowcroft J, Yoneki E (2008) Bubble rap: social-based forwarding in delay tolerant networks. In: Proceedings of ACM MobiHoc, pp 241–250
- Socievole A, De Rango F (2015) Energy-aware centrality for information forwarding in mobile social opportunistic networks. In: IEEE IWCMC. IEEE, pp 622–627
- Zhu Y, Zhang C, Mao X, Wang Y (2015) Social based throwbox placement schemes for large-scale mobile social delay tolerant networks. *Comput Commun* 65:10–26
- Chaintreau A, Hui P, Crowcroft J, Diot C, Gass R, Scott J (2007) Impact of human mobility on opportunistic forwarding algorithms. *IEEE Trans Mob Comput* 6(6):606–620
- Karagiannis T, Le Boudec J, Vojnović M (2010) Power law and exponential decay of intercontact times between mobile devices. *IEEE Trans Mob Comput* 9(10):1377–1390
- Passarella A, Conti M (2011) Characterising aggregate inter-contact times in heterogeneous opportunistic networks. In: NETWORKING 2011. Springer, Berlin, pp 301–313
- Tang J, Kim H, Anderson R (2012) Temporal node centrality in complex networks. *Phys Rev E* 85(2):026107
- Kim H, Tang J, Anderson R, Mascolo C (2012) Centrality prediction in dynamic human contact networks. *Comput Netw* 56(3):983–996
- Zhou H, Xu S, Huang C (2015) Temporal centrality prediction in opportunistic mobile social networks. In: Hsu C-H, Xia F, Liu X, Wang S (eds) Internet of vehicles-safe and intelligent mobility. Springer, Berlin, pp 68–77
- Huang D, Zhang S, Hui P, Chen Z (2015) Link pattern prediction in opportunistic networks with kernel regression. In: International conference on communication systems and networks, IEEE, pp 1–8
- Wei W, Carley K (2015) Measuring temporal patterns in dynamic social networks. *ACM Trans Knowl Discov Data* 10(1):9
- Scott J, Gass R, Crowcroft J, Hui P, Diot C, Chaintreau A (2009) Crawdad data set cambridge/haggle (v. 2009-05-29)
- Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. *Proc Natl Acad Sci* 106(36):15274–15278. doi:10.1073/pnas.0900282106