Centrality prediction based on *K*-order Markov chain in Mobile Social Networks

Mengni Ruan[1] · Xin Chen[1] · Huan Zhou[1]

Abstract

In this paper, we proposed a centrality prediction method based on K-order Markov chains to solve the problem of centrality prediction in Mobile Social Networks (MSNs). First, we use the information entropy to analyze the past and future regularity of the nodes' centrality in the real mobility traces, and verify that nodes' centrality is predictable. Then, using the historical information of the center of the node, the state probability matrix is constructed to predict the future central value of the node. At last, through the analysis of the error between real value and predicted value, we evaluate the performance of the proposed prediction methods. The results show that, when the order number is $K = 2$, compared with other existing four time-order-based centrality prediction methods, the proposed centrality prediction method based on K-order Markov chain performs much better, not only in the MIT Reality trace, but also in the Infocom 06 traces.

1 Introduction

With the rapid development of mobile smart devices (such as ipads, PDAs, and smart-phones), Mobile Social Networks (MSNs) have began to emerge in our daily life [1–3]. In MSNs, mobile smart devices can be considered as nodes, and short-distance communications between devices can be viewed as edges between nodes, which often appear or disappear over time. Since carriers of mobile smart devices are individual members of society, data is mostly spread among nodes with social relations, such as friends, classmates, and family members [4–9]. Although Mobile

Social Network is inherently a dynamically connected network with time-varying topology, the activity of nodes in the network with this characteristic is not irregular. Actually, the mobility of nodes in MSNs mainly depends on human behavior patterns. Some studies have showed that the activities of individuals or groups of people are generally characterized by regularity, aggregation and social characteristics, and the regularity of human intrinsic activity makes the predictability of people's behavior as high as 93% [10].

In recent years, more and more research has used social network analysis technology to help design routing protocols [11–15]. Centrality is one of the focuses of social network analysis. The greater the influence, the more likely a node is to contact other nodes [16–19]. Therefore, this paper will use node's centrality to analyze the social characteristics of MSNs. Previous studies have proposed some centrality metrics to measure the importance of nodes such as betweenness centrality, closeness centrality, degree centrality, and so on [20–22]. At present, some studies in MSNs have tried to predict the centrality (or importance) of nodes in the future. For example, Kim et al. in [23] have proposed several methods to predict nodes' future importance under three important centrality metrics, namely degree, closeness, and betweenness centrality. However, according to the evaluation results, the proposed methods have a large difference in terms of performance. The main

✉ Huan Zhou
  zhouhuan117@gmail.com

  Mengni Ruan
  mnruan@163.com

  Xin Chen
  sexychenxin@gmail.com

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

reason is that the proposed methods fail to make full use of discrete time series information. To solve this problem, in the paper we propose a prediction method based on the Markov chain. The Markov chain is a theory that studies the state of things and their transition probabilities, which can not only be applied to time and space sequences, but also to represent randomness. Because of the continuous nature of the development of things, which makes a very close relationship between adjacent variables, there is a certain limitation in the 1-order Markov chain. In order to make the Markov chain not only represent the real situation, but also improve the prediction performance, Bartlett [24] proposed the concept of high-order Markov chain, which extends the traditional 1-order Markov chain correlation to high-order correlation. In MSNs, because of its own social nature, nodes will have a certain continuity. That is, the previous period has a significant impact on the latter period, which is similar to the nature of Markov chain. However, when using the 1-order Markov chain model to predict the state of the future, a lot of historical state information is ignored and only the information of the current moment is used. Therefore, this paper proposes a prediction method based on K-order Markov chain to predict the future importance of nodes in MSNs. The main contributions of this paper are as follows:

1. By analyzing the information entropy of the node's centrality, we found the regularity and relativity of the node centrality.
2. The K-order Markov chain-based prediction model is proposed, and extensive simulations are conducted to determine the optimal order K.
3. The K-order Markov chain-based prediction method is compared with other four existing prediction methods. The results show that the K-order Markov chain-based prediction method has great advantages.

The remainder of this paper is organized as follows. Section 2 gives a brief introduction about the related work. Section 3 introduces the K-order Markov chain and three centrality metrics. Section 4 uses the information entropy to analyze the past and future regularity of the nodes' centrality. Section 5 introduces K-order state transition matrix and prediction model. Extensive simulations are conducted to evaluate the performance of the proposed methods in Section 6. Section 7 concludes the paper.

2 Related work

Human daily activities are regular. For groups, such as school classes, teams and interest groups, the frequency and time of their behavior and the law of group activities are relatively stable [25–27]. People often say that "things

are gathered together, people are divided into groups" and nodes with the same or similar nature will come together to form small groups [28, 29]. The small world characteristics of human mobility behavior makes human society exhibit high clustering characteristics, and a few people occupy the core position of social networks [30]. The social nature of people's behavior means that people are more inclined to move around in places they are relatively familiar with, and more likely to choose to be with people who are familiar with them [31]. This shows that the centrality of the temporal state of the nodes in MSNs is predictable.

Centrality is a measure of the influence of a node in a network. For simplicity, most studies in MSNs tend to model MSNs as the static network to analyze nodes' centrality [16]. For example, authors in [32] proposed SimBet Routing which uses the betweenness centrality metric and locally social similarity to improve data transmission efficiency. Authors in [23] proposed a method of sequential graph, which transforms the topological structure of the network into a set of static graphs based on time sequence. Authors in [33] proposed a basic theoretical framework for sequential graph modeling. In sequential graph modeling, MSNs are modeled as a set of continuous graphs in a continuous period of time. The structure of the MSNs in each time period is considered to be basically not change. Authors in [34] consider the dynamic network of MSNs as a set of network topology snapshots from a series of minimum time units. In this model, network characteristics of MSNs' connectivity, sparsity, and data forwarding indicators are considered. Authors in [35] developed a more general model by introducing a variable representing the speed at which a message travels. Furthermore, authors in [36] also proposed temporal centrality metrics based on temporal paths in order to measure the importance of a node in a dynamic network.

To measure nodes' centrality more accurately, some studies have tried to model MSNs as a time-varying network, and propose several methods to predict nodes' future centrality based on time-varying graphs. For example, authors in [23] have proposed several methods to predict nodes' future importance under three important centrality metrics, namely degree, closeness, and betweenness centrality. Similarly, through extensive real trace-driven simulations, authors in [7] observed that nodes' temporal centrality shows strong correlation. With this knowledge, they designed several intuitive methods to predict nodes' future temporal centrality. However, the results also showed that the performance of different prediction methods are obviously different. The main reason is that the proposed prediction methods fail to make full use of discrete time series information, while the Markov process is a theory that studies the state of things and their transition probabilities [24]. Authors in [37] used Markov chains to predict the sociality of future vehicles and proposed two greedy heuristics to

1664

Peer-to-Peer Netw. Appl. (2019) 12:1662–1672

select the most "central" vehicles as seeds for mobile advertising. Based on the above work, our work tries to use the Markov chain model to predict nodes' centrality in MSNs.

3 Basic knowledge

In this section, we first introduce the K-order Markov chain model used in this paper and then present the centrality metrics which are used to analysis information entropy of nodes and prediction problem.

3.1 K-order Markov chain

The Markov chain describes a state sequence, which is described in the mathematical field as a discrete time random variable with Markov properties. The main idea is that given the state information of the current moment, the past (i.e. the historical state of the present moment) is irrelevant with the predictions of the future (i.e. the future state of the present moment), that is, the past and the future are independent of each other. When there are n consecutively changing things, and in the course of its change, the result of any change is non-responsive, then the set of these continuous changes is called the Markov chain, and the process of evolution of such things is called Markov process [38–41].

Because the development of things is continuous, there is a high correlation between the state of the adjacent neighbor. However, when using the 1-order Markov chain model to predict the state of the future, a lot of historical state information is ignored and only the information of the current moment is used. Therefore, such limitations make the practical application of 1-order Markov chain prediction method lack of prediction accuracy. In order to improve the prediction accuracy, authors in [42] use K-order Markov chain model to estimate network link packet loss in time domain. In this paper, similarly, we use the K-order Markov chain model to predict nodes' future centrality in MSNs. The K-order Markov chain model extends the order of Markov chain in the observation sequence from 1-order correlation to K-order correlation, the advantage of which is that it can make better use of historical information and establish a reasonable and very close relation between several neighboring historical states, present moment states and future states. It can be defined as:

$$
\begin{aligned}
P\left(C_t = i_t | C_{t-1} = i_{t-1}, ..., C_0 = i_0\right) = \\
P\left(C_t = i_t | C_{t-1} = i_{t-1}, ..., C_{t-k} = i_{t-k}\right)
\end{aligned}
\tag{1}
$$

3.2 Centrality metrics

The importance of nodes is the most important research hotspot in MSNs, which refers to the influence of nodes.

From the point of view of the network topology, the importance of a node is not its independent property, but the common relationship that the node encounters with other nodes. The network centrality of a node can measure the importance of nodes in a network topology. Centrality metrics is a very important concept in network analysis, which is used to measure the importance of nodes in the network. There are several common methods to measure "centrality". In this paper, we only introduce three of them: degree, betweenness and closeness centrality. Formally, we use the standard definition of the degree, betweenness and closeness centrality, and the centrality value of a node i can be expressed as follows [13].

3.2.1 Degree centrality

Degree centrality represents the total number of direct links with other nodes of a certain node. Higher degree centrality value of a node means more contacts with other nodes in the network, and the Degree centrality value of a certain node i is expressed as:

$$
Degree(i) = \frac{\sum\limits_{j \neq i, j \in V} e(i, j)}{|V| - 1}
\tag{2}
$$

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.

3.2.2 Betweenness centrality

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

$$
Betweenness(i) = \sum\limits_{u \neq i, v \neq i, i \in V} \frac{\delta_{u,v}(i)}{\delta_{u,v}}
\tag{3}
$$

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

3.2.3 Closeness centrality

Closeness centrality represents the distance a certain node to all other reachable nodes in the network, which can be calculated as the average shortest path length between

a certain node and all other reachable nodes. Closeness centrality value of a certain node i is expressed as:

$$Closeness(i) = \frac{1}{|V| - 1} \sum_{j \neq i, j \in V} \Delta_{i,j}(i) \qquad (4)$$

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 Information entropy analysis

In this section, the information entropy is used to describe the change rule of node centrality in MSNs. The concept of entropy comes first from thermodynamics in physics, but entropy in information theory has nothing to do with the entropy of thermodynamics. In information theory, one measure of uncertainty is the use of entropy. Information entropy is an analysis of the uncertainty of objective outcome in the angle of random experiment, which is used to calculate the expected value of random variable, the greater the information entropy of this random variable, the greater the uncertainty of the variable. Therefore, information entropy can be used to measure the regularity of the system, and when the regularity of a system is higher, the information entropy will be smaller.

C represents the degree centrality random variable of the node. Assuming the size of the observation sequence is S, then this observation sequence can be expressed as a vector $V = (v_0, v_1, ..., v_{S-1})$, where $0 \leq i \leq S - 1$, v_i represents the degree of centrality of the node in the i-th window. The probability of the value j is c_j/S, where c_j represents the number of times the value is j. Therefore, the entropy of V can be expressed as:

$$E(C) = \sum_{j=0}^{\infty} (c_j/S) \log_2 \frac{1}{c_j/S} \qquad (5)$$

Here, when $K = 1$, for a given node's current degree centrality C, the random variable C' represents the degree centrality of the previous time window of the node. If S is large enough, C' and C have the same distribution. Vector V can be expressed as $W = \{(v_i, v_{i+1}) : 0 \leq i \leq S - 2\}$. Therefore, the joint entropy of C' and C can be expressed as:

$$E(C', C) = \sum_{(c', c) \in W} p(c', c) \log_2 \frac{1}{p(c', c)} \qquad (6)$$

where $p(c', c)$ is the number of occurrences of (c', c) in W divided by the total number of pairs of elements in W.

When $E(C)$ and $E(C', C)$ are known, the conditional entropy of C for a given C' is:

$$E(C|C') = E(C', C) - E(C')$$
$$= E(C', C) - E(C) \qquad (7)$$

When $K = 2$, for a given node's current degree centrality C, the random variable C' represents the degree centrality of the first two window of the node. Similarly, we can get conditional entropy $E(C|C')$ as follows:

$$E(C|C'') = E(C'', C) - E(C'')$$
$$= E(C'', C) - E(C', C) \qquad (8)$$

The centrality sequence of each node is constructed with the collected data set, and the marginal entropy and conditional entropy of the current state of the node are computed when the centrality of the K time window is given ($K = 0$ is the marginal entropy of the current state of the node, and $K > 0$ is the conditional entropy of a node where the K state before the current state is known). Figure 1 gives the simulation results of the marginal entropy
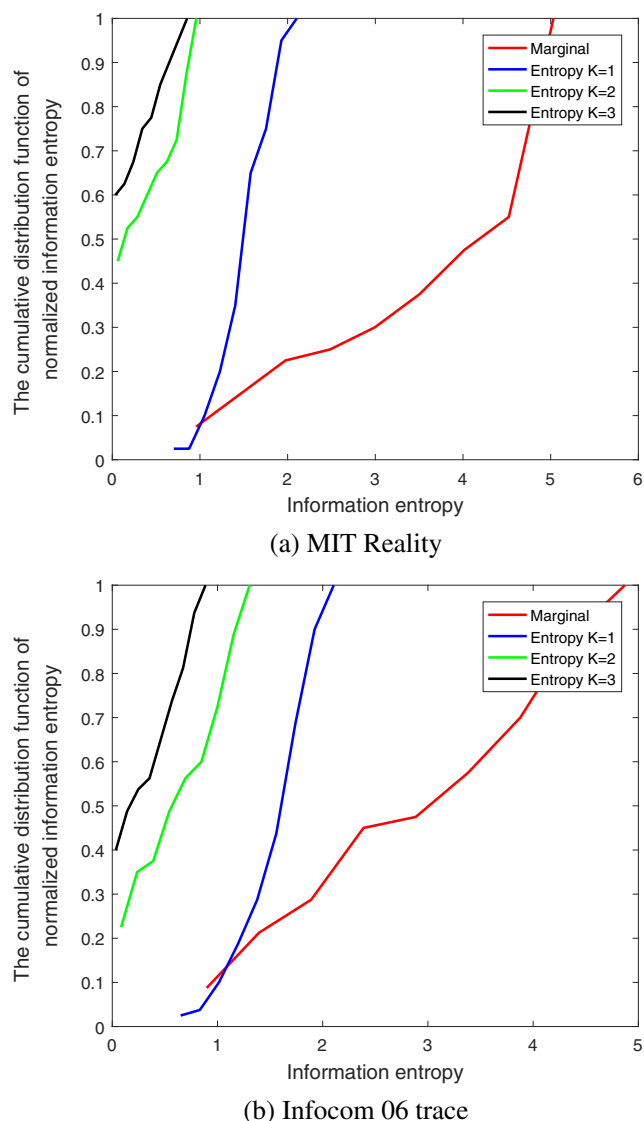

(a) MIT Reality

(b) Infocom 06 trace

Fig. 1 Cumulative distribution function of the entropy of the degree centrality

1666

Peer-to-Peer Netw. Appl. (2019) 12:1662–1672

and conditional entropy of the degree centrality of the node in the MIT Reality and the Infocom 06 traces.

In Fig. 1, where the horizontal axis represents information entropy, the longitudinal axis represents the cumulative distribution function of normalized information entropy. In the case of $K = 0, 1, 2, 3$, the cumulative distribution of the marginal entropy and conditional entropy of the node in the MIT Reality trace and the Infocom 06 trace is shown in Fig. 1 ($K = 0$ is the marginal entropy of the current state of the node. When $K > 0$ is the conditional entropy of a node where the K state before the current state is known). It can be seen from the figure that the conditional entropy of the $K = 1$ is much smaller than the marginal entropy, and the conditional entropy of the $K = 2$ is much smaller than the conditional entropy of the $K = 1$. At the same time, the conditional entropy is obviously less than the marginal entropy, and the conditional entropy decreases with the increase of K value. This phenomenon indicates that the degree centrality of the node has certain regularity, and the uncertainty of the degree centrality of the node decreases with the knowledge of some degree centrality information before the present, which provides a possibility for the prediction of the centrality of the node's future.

Figure 2 shows the comparison of the marginal mean entropy and the conditional mean entropy of degree centrality, betweenness centrality and closeness centrality of the nodes in the MIT Reality and the Infocom 06 traces. It can be obtained from Fig. 2 that the conditional mean entropy is obviously less than the marginal mean entropy. And the conditional mean entropy is obviously reduced as the value increases. This conclusion is similar to the conclusion drawn from Fig. 1, which proves the regularity and relativity of node centrality, and provides a theoretical basis for predicting the future centrality of nodes based on K-order Markov chain method.

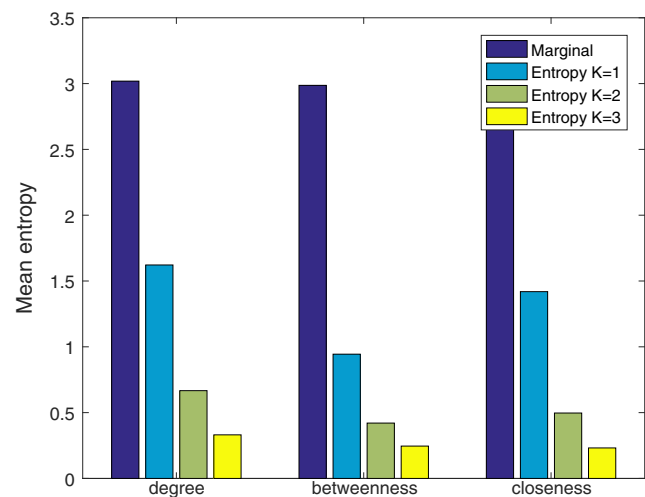## 5 Prediction based on K-order Markov chain

In this section, we introduce the K-order state transition matrix and construct the prediction model which based on the K-order Markov chain prediction method.

5.1 K-Order state transition matrix

Based on the K-order Markov chain prediction method, the prediction of node's centrality is the probability of calculating all possible states by the current known historical state information, in which the maximum probability state is the desired state, that is the prediction centrality value. The state transition probability matrix is composed of the probability of state, and the centrality
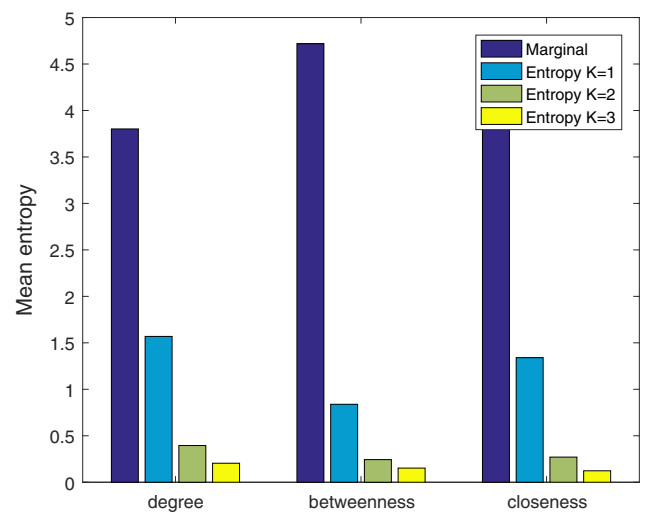
(a) MIT Reality

(b) Infocom 06 trace

Fig. 2 Mean entropy of different centrality metrics

prediction method based on K-order Markov chain is mainly to solve the state transition probability matrix.

When solving the K-order state probability transfer matrix S, the centrality of nodes in the next time window is predicted by the centrality of the adjacent k time window of the current time window. In the state transition probability matrix S, the matrix element represents the probability of arriving at state b from a state a, after a k time unit, and becomes a k-step state transition probability:

$$m_{a,b} = P\left(C_{n+1} = b | C\left(n - k + 1, n\right) = a\right) \qquad (9)$$

where v represents any continuous k centrality value in $(v_0, v_1, ..., v_{S-1})$, and b represents an independent centrality value.

The solution of the state transition matrix S is obtained by a large number of centrality historical data of the node, and as same with the information entropy, it can approximate its probability by the frequency at which the nodal centrality value appears:

$$P\left(C_{n+1} = b|a\right) = \frac{N\left(b, a\right)}{N\left(v\right)} \tag{10}$$

where $N(v)$ represents the number of consecutive occurrence of the centrality value of k in a certain state as a in the historical data sequence, and $N(b, a)$ represents the number of times the next state is b after the state is a in the historical data sequence.

5.2 Prediction model

Given a time series contact graph, for each node v_i, calculating the centrality of each node in the contact graph, a series of centrality values can be obtained, denoted as $\{x_i\}_{i=1}^n$. By discretizing the continuous measures, a finite state space can be obtained, expressed as S. The Markov chain of k-state transition probabilities can be used to estimate for all a $\in$ S and $\underline{b} \in S^k$, where $\underline{b} = (b_1, b_2, ..., b_k)$. Remember the $n_{\underline{b}a}$ is the number of times the state b follows the value a in the sequence. Note the $n_{\underline{b}}$ is the number of occurrences of state $\underline{b}$, and $p_{\underline{b};a}$ represents the estimation of the state transition probability from state $\underline{b}$ to state $(b_2, ..., b_k, a)$. The state transition probability of the maximum likelihood estimator of the K-order Markov chain is:

$$p_{\underline{b};a} = \begin{cases} n_{\underline{b}a}/n_{\underline{b}} , & if \ n_{\underline{b}} > 0 \\ 0 , & otherwise \end{cases} \tag{11}$$

Specifically, $\underline{b}_i$ represents the current state of node v_i in the K-order Markov chain. The centrality of node v_i in the next window can be calculated as:

$$C_{fu}^i = \sum_{a \in S} p_{\underline{b};a} \times a \tag{12}$$

In the process of predicting node's future centrality, two parameters are critical to the computational accuracy: the order of Markov models and the historical data length used for training models. For Markov chain models with known state sets, simply increasing k does not necessarily apply to the inclusion of dependencies in time series. The order of Markov chains can be evaluated by information content test [43].

6 Performance evaluation

In order to evaluate the performance and accuracy of the proposed prediction methods, extensive simulation experiments are carried out in real traces. Here, we also use two real mobility traces, the MIT Reality and Infocom 06 traces to do extensive simulations. For each prediction method, we use the average error between the predicted value and the real value to analyze the prediction precision of the proposed prediction method, and we compare the predicted results with the performance of several prediction methods based on time windows.

6.1 Simulation setup

In the process of data processing, because the centrality value of the node retention accuracy is high, the difference between the values is small. If the value of each centrality is an independent state, it will not only cause the states to be more, but also cause the state to be difficult to match. At the same time, when the state transition probability matrix is computed, the calculation of time and space is more complicated. Therefore, it is necessary to classify the nodes according to their centrality value. In this chapter, for the calculation of the state transition probability matrix, the probability of each state is calculated with the current state as the starting state. Take the MIT dataset closeness centrality of the node as the example to illustrate.

There are 97 nodes in the MIT Reality, and a total of 77 different closeness centrality values are obtained by computational statistics. After sorting, you can find that due to the high accuracy of data retention, the number of effective digits after the decimal point is not large, so the difference between the closeness centrality value is not big, but if the 77 different closeness centrality value as a single independent state to calculate, it will cause the states of node to be too big, the state transition probability matrix is large and inconvenient to calculate, it can also make the states difficult to match. Therefore, the closeness centrality value is artificially divided. In the experiment process, the closeness centrality value is divided into a state with little difference, which can reduce the total number of states in the calculation process, so as to simplify the calculation.

The main performance index of the evaluation forecasting method is its prediction precision, that is, the size of the prediction error. The prediction error is the difference between the predicted value and the real value, which can be divided into absolute error and relative error according to different characteristics, and the absolute error is also called the position deviation, and the relative error is also called the scale deviation. The absolute error reflects the size of the measured deviation from the real value, where the error analysis of the centrality predictive method based on Markov is carried out with absolute error. |V| is the number of nodes in the network, you can define the prediction method error as the average of the predicted errors for all nodes in the network:

$$\text{Error}(G_{r+1}) = \frac{\sum_{u \in V} \left| C_{r+1}(u) - \hat{C}_{r+1}(u) \right|}{|V|} \tag{13}$$

1668

Peer-to-Peer Netw. Appl. (2019) 12:1662–1672

For each prediction method, the error $Error(G_{r+1})$ between the predicted value and the real value is used to analyze the prediction accuracy of the proposed prediction method.

6.2 Performance comparison

It is necessary to solve the problem of state space expansion based on the proposed centrality prediction method. The K-order state transition probability matrix is constructed using nodes' history centrality scale of N, and the state probability transfer matrix's size is $N^k \times N$. If the state number N and the order K values are very large, the state transition probability matrix will be large, and the complexity of computation and search is very large. It has been proved

(a) Degree centrality in the MIT Reality trace

(b) Degree centrality in the Infocom 06 trace

Fig. 3 Prediction of degree centrality using the K-order Markov chain-based prediction method

in the relevant literatures that low order Markov chains can also achieve good prediction results in the case of large data volume [44]. Therefore, this paper makes a judgment on the order of K in Markov chain by simulation experiment on the basis of the collected real mobility traces.

Figure 3 shows the prediction results of the degree centrality values in the MIT Reality and Infocom 06 traces using a centrality prediction method based on $K = 1, 2, 3$-order Markov chain, where the horizontal axis represents the number of time windows used for the predicted data, and the longitudinal axis represents the predicted error between the predicted value and the real value. The results show that when the order $K = 1$, the prediction error is big and the fluctuation and volatility are large, the performance is not stable and the prediction effect is not good. When the order $K = 2$, the prediction error reaches the minimum value, and the volatility is small and the performance is stable. When the order $K = 3$, the performance of Markov chain prediction method compared to $K = 2$, although the prediction error is relatively increased, but the fluctuation is less and the prediction result is relatively stable. With the growing of K in K-order Markov chain, the complexity of computation increases exponentially, so the computational complexity of prediction method and the prediction error are considered synthetically, based on K-order Markov chain prediction method, when $K = 2$ the prediction method based on Markov chain has the least error and relatively good stability in each dataset.

We summarize the above results of the simulation experiments using the K-order Markov chain-based centrality prediction method in the MIT Reality and Infocom 06 traces. With the increase of the order K, the time complexity and the spatial complexity of the computational scale will increase exponentially, although the Markov chain-based prediction method of order $K = 2$ is obviously better than that of $K = 1$, the prediction performance of order $K = 3$ Markov chain-based prediction method is not better than the order $K = 2$ in terms of prediction error and stability. One aspect of this result is that the prediction method based on the K-Markov chain is only a probabilistic prediction, and its prediction results can only indicate that the system is approaching a certain state in the future with a certain probability, rather than absolutely infinite approximation to this state. On the other hand, when calculating the state transition probability matrix, the artificial manual state division is adopted according to the value of the node's centrality. Therefore, comparing to the $K = 2$ Markov chain-based prediction method, the prediction performance of the Markov chain-based prediction method of order $K > 2$ in the simulation experiment does not have great improvement. Therefore, in the following, our proposed K-order Markov chain-based centrality prediction method in MSNs is modeled using the 2-order Markov chain.
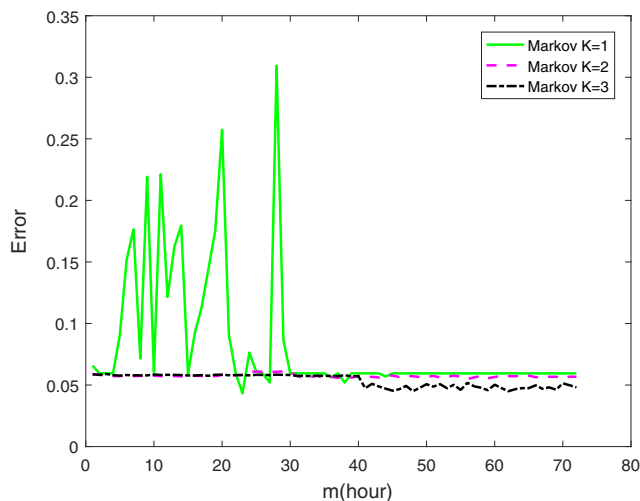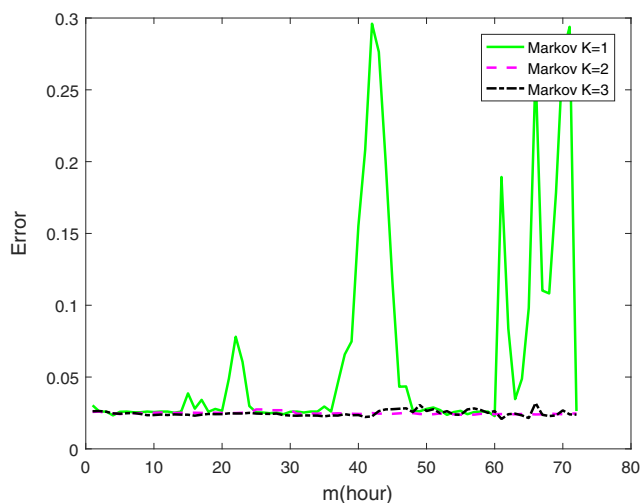
After choosing the optimal K of the K-order Markov chain-based prediction method, we compare our proposed prediction method with four other existing prediction methods introduced in [7]:

1. Last Method: Only using the node's centrality value in the last observation time window as the centrality value of the node in the next window.
2. Recent Uniform Average Method: Using the average node's centrality value of the last m observation windows to predict the centrality value of the node in the next window.
3. Recent Weighted Average Method: Using the weighted average node's centrality value of the last m observation windows to substitute the uniform average centrality value.
4. Periodical Average Method: For people's daily lives, the reasonable period is one day or one week. For periods of one day or one week, consider using the period-averaged centrality value as the centrality value of the future prediction window.

Figure 4 shows the centrality prediction results for different prediction methods in the MIT Reality. It can be seen from the figure that compared with four other existing centrality prediction methods, the performance of our proposed prediction method based on 2-order Markov chain is obviously better when predicting three centrality metrics. Our proposed centrality prediction method not only in the prediction error, but also in the stability of the prediction error has a greater advantage. When compared with the best performance recent weighted average method, the prediction errors are not much different in the prediction of degree centrality, and their prediction errors are unbiased in the prediction of the betweenness centrality and closeness centrality. At the same time, it can be seen that the prediction method based on the 2-order Markov chain has less volatility.

Figure 5 shows the centrality predictive results for different prediction methods in the Infocom 06 traces. It can be seen from the figure that compared with the other four existing centrality prediction methods, the performance of our proposed centrality prediction method is obviously better. When compared with the best-performance last method, their prediction errors are basically the same when predicting the degree centrality, while for the prediction error of the betweenness centrality and closeness centrality, the prediction performance of our proposed centrality prediction method is much better. Therefore, our proposed 2-order Markov chain-based prediction method is superior to other existing prediction methods.
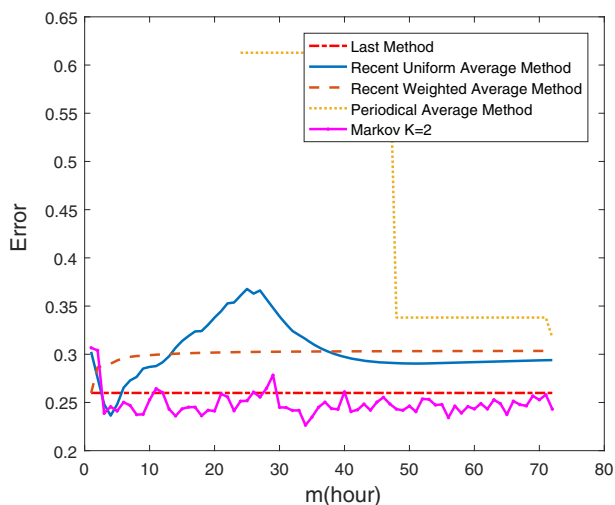
Therefore, the above results prove that our proposed prediction method is superior to other existing prediction methods, can be applicable to both MIT Reality and Infocom 06 traces.
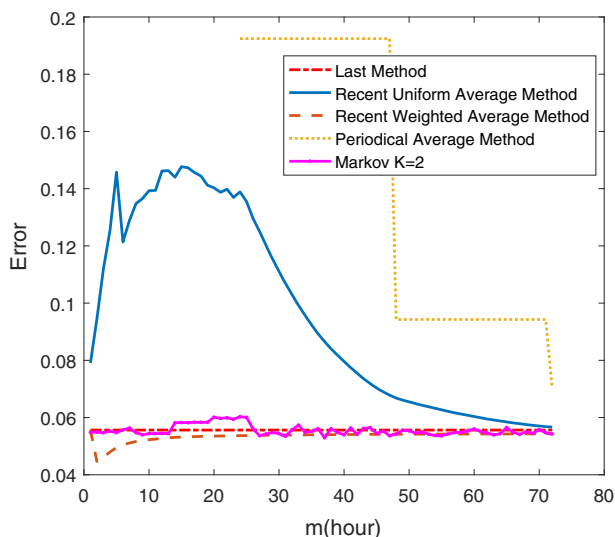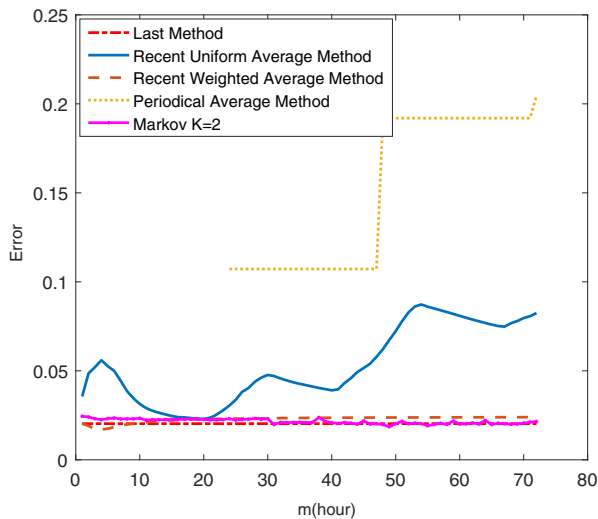

(a) Degree centrality
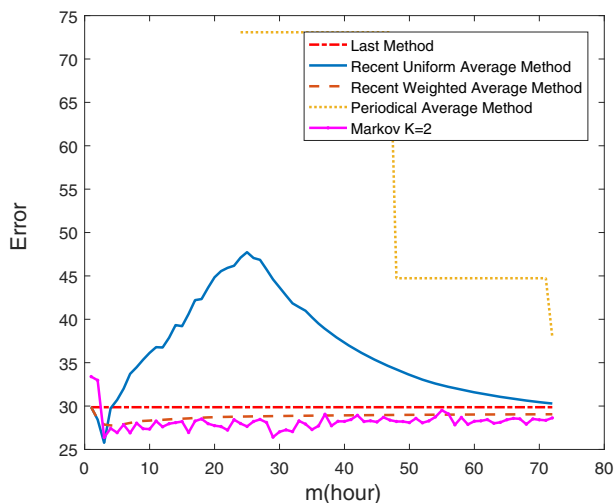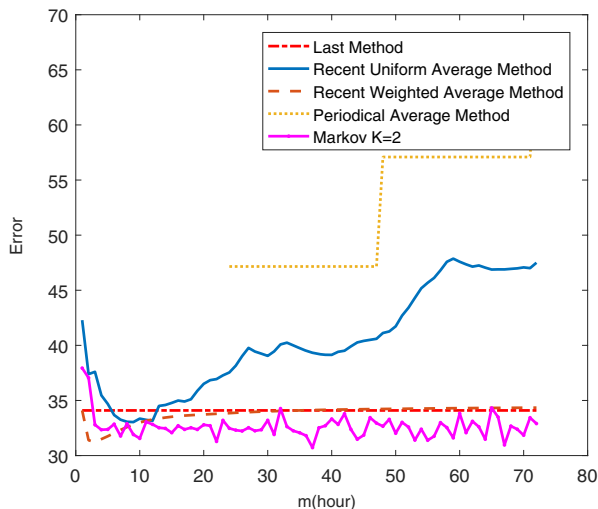

(b) Betweenness centrality
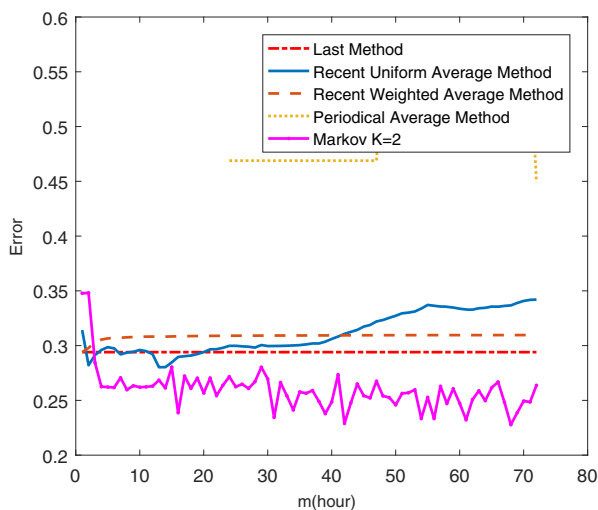

(c) Closeness centrality

Fig. 4 Comparison of prediction methods in the MIT Reality trace

(a) Degree centrality

(b) Betweenness centrality

(c) Closeness centrality

Fig. 5 Comparison of prediction methods in the Infocom 06 trace

7 Conclusions

This paper analyzed the centrality of nodes in MSNs and proposed an effective method for predicting nodes' future centrality. By analyzing the information entropy of a large number of nodes' centrality values using the real mobility traces, it is proved that when the centrality of the node is known, the uncertainty of the centrality of the node is reduced, providing a theoretical basis for predicting the future centrality of the node. Discovering the limitation of 1-order Markov chain model, this paper proposed nodes' centrality prediction method using historical information based on K-order Markov chain, and gave the conclusion that the order of K in Markov chain is based on the analysis of the results by experiment comparison. Finally, extensive real trace-driven simulations are conducted to evaluate the performance of our proposed K-order Markov chain-based centrality prediction method, using three centrality metrics, namely betweenness centrality, closeness centrality, and degree centrality. The results show that compared with other existing prediction methods, our proposed prediction method performs much better not only in the MIT Reality trace, but also in the Infocom 06 trace.

References

1. Xiao M, Wu J, Huang L (2014) Community-aware opportunistic routing in mobile social networks. IEEE Trans Comput 63(7):1682–1695
2. 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
3. Zhou H, Wang H, Li X, Leung VCM (2018) A survey on mobile data offloading technologies. IEEE Access 6(1):5101–5111
4. Yang G, He S, Shi Z (2017) Leveraging crowdsourcing for efficient malicious users detection in large-scale social networks. IEEE Internet Things J 4(2):330–339
5. 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
6. Yang G, He S, Shi Z, Chen J (2017) Promoting cooperation by social incentive mechanism in mobile crowdsensing. IEEE Commun Mag 55(3):86–92
7. Zhou H, Leung VCM, Zhu C, Xu S, Fan J (2017) Predicting temporal social contact patterns for data forwarding in opportunistic mobile networks. IEEE Trans Veh Technol 66(11):10372–10383
8. Zhu C, Leung VCM, Rodrigues JJPC, Shu L, Wang L, Zhou H (2018) Social sensor cloud: framework, greenness, issues, and outlook. IEEE Netw 32(5):100–105
9. Zhu C, Zhou H, Leung VCM, Wang K, Zhang Y, Yang LT (2017) Toward big data in green city. IEEE Commun Mag 55(11):14–18
10. Galton F (1886) Regression towards mediocrity in hereditary stature. J Anthropol Inst G B Irel 1(15):246–263

Peer-to-Peer Netw. Appl. (2019) 12:1662–1672

1671

11. 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

12. Yuan Q, Cardei I, Wu J (2009) Predict and relay: an efficient routing in disruption-tolerant networks. In: ACM interational symposium on mobile ad hoc networking and computing, MOBIHOC 2009, New Orleans, La, Usa, pp 95–104

13. Chen H, Lou W (2014) Gar: group aware cooperative routing protocol for resource-constraint opportunistic networks. Comput Commun 48(18):20–29

14. Zhou H, Ruan M, Zhu C, Leung VCM, Xu S, Huang C (2018) A time-ordered aggregation model-based centrality metric for mobile social networks. IEEE Access 6(1):25588–25599

15. Huang C, Chen Y, Xu S, Zhou H (2018) The vehicular social network (vsn)-based sharing of downloaded geo data using the credit-based clustering scheme. IEEE Access 6(1):58254–58271

16. Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Networks 1(3):215–239

17. Pan H, Chaintreau A, Scott J, Gass R, Crowcroft J (2005) C.diot, Pocket switched networks and human mobility in conference environments. In: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking, pp 244–251

18. Dong ZB, Song GJ, Xie KQ, Wang JY (2009) An experimental study of large-scale mobile social network. In: Proceedings of the 18th international conference on World Wide Web, pp 1175–1176

19. Zhou H, Tong L, Xu S, Huang C, Fan J (2016) Predicting temporal centrality in opportunistic mobile social networks based on social behavior of people. Pers Ubiquit Comput 20(6):1–13

20. Scott J (1988) Trend report social network analysis. Sociology 22(1):109–127

21. Gao W, Li Q, Zhao B, Cao G (2009) Multicasting in delay tolerant networks: a social network perspective. In: Proceedings of the 10th ACM international symposium on Mobile ad hoc networking and computing, pp 299–308

22. Wang S, Huang L, Hsu CH, Yang F (2016) Collaboration reputation for trustworthy web service selection in social networks. J Comput Syst Sci 82(1):130–143

23. Kim H, Tang J, Anderson R, Mascolo C (2012) Centrality prediction in dynamic human contact networks. Comput Netw 56(3):983–996

24. Bartlett MS (1951) The frequency goodness of fit test for probability chains. Math Proc Camb Philos Soc 47(1):86–95

25. Nguyen HA, Giordano S (2012) Context information prediction for social-based routing in opportunistic networks. Ad Hoc Netw 10(8):1557–1569

26. Bulut E, Szymanski BK (2012) Exploiting friendship relations for efficient routing in mobile social networks, vol 23

27. Zhou H, Wang H, Chen X, Li X, Xu S (2018) Data offloading techniques through vehicular ad hoc networks: a survey. IEEE Access 6(1):65250–65259

28. Li F, Zhao L, Zhang C, Gao Z, Wang Y (2014) Routing with multi-level cross-community social groups in mobile opportunistic networks. Pers Ubiquit Comput 18(2):385–396

29. Williams MJ, Whitaker RM, Allen SM (2012) Decentralised detection of periodic encounter communities in opportunistic networks. Ad Hoc Netw 10(8):1544–1556

30. Hossmann T, Spyropoulos T, Legendre F (2011) A complex network analysis of human mobility. In: Proceedings of the IEEE Conference on Computer Communications Workshops(INFOCOM WKSHPS), pp 876–881

31. Wang D, Pedreschi D, Song C (2011) Human mobility, social ties, and link prediction. In: ACM SIGKDD international conference on knowledge discovery and data mining, pp 1100–1108

32. Daly E, Haahr M (2007) Social network analysis for routing in disconnected delay-tolerant manets. In: ACM Interational symposium on mobile Ad Hoc networking and computing, pp 32–40

33. Kempe D, Kleinberg J, Kumar A (2002) Connectivity and inference problems for temporal networks. J Comput Syst Sci Int 64(4):820–842

34. Ferreira A (2004) Building a reference combinatorial model for manets. IEEE Netw 18(5):24–29

35. Tang J, Musolesi M, Mascolo C, Latora V (2010) Characterising temporal distance and reachability in mobile and online social networks. ACM SIGCOMM Comput Commun Rev 40(1):118–124

36. Tang J, Musolesi M, Mascolo C, Latora V, Nicosia V (2010) Analysing information flows and key mediators through temporal centrality metrics. In: Proceedings of the 3rd workshop on social network systems, pp 1–6

37. Qin J, Zhu H, Zhu Y, Lu L, Xue G, Li M (2014) Post: Exploiting dynamic sociality for mobile advertising in vehicular networks. IEEE Trans Parallel Distrib Syst 27(6):1–1

38. Saruukkai R (2000) Ramesh link prediction and path analysis using Markov chains. Comput Netw 33(1):377–386

39. Zhu Y, Zhong Z, Zheng WX, Zhou D (2018) Hmm-based h-infinity filtering for discrete-time Markov jump lpv systems over unreliable communication channels, IEEE Transactions on Systems. Man, and Cybernetics: Systems 48(12):2035–2046

40. Zhu Y, Zhang L, Zheng WX (2016) Distributed h-infinity filtering for a class of discrete-time Markov jump lur' systems with redundant channels. IEEE Transactions on Industrial Electronics 63(3):1876–1885

41. Zhang H, Qi Y, Zhou H, Zhang J, Sun J (2017) Testing and defending methods against dos attack in state estimation. Asian J Control 19(3):1–11

42. Fei GL (2011) Unicast network loss tomography based on k-th order Markov chain. J Electron Inf Technol 33(9):2278–2282

43. Burnham KP, Anderson DR (2004) Multimodel inference understanding aic and bic in model selection. Sociol Methods Res 33(33):261–304

44. Song L, Kotz D, Jain R, he X (2006) Evaluating next-cell predictors with extensive wi-fi mobility data. IEEE Trans Mob Comput 5(12):1633–1649
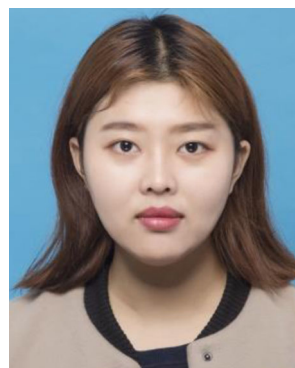
Mengni Ruan received the bachelor's degree from the Hubei University of Arts and Science. She is currently a Graduate Student with the College of Computer Information and Technology, China Three Gorges University. Her main research interests are mobile data offloading and mobile social networks.

Xin Chen received his B.S. Degree from China Three Gorges University. Currently, he is a graduate student at the College of Computer Information and Technology, China Three Gorges University. His main research interests are mobile data offloading and VANETs.

Huan Zhou received his Ph.D degree from the Department of Control Science and Engineering at Zhejiang University. He was a visiting scholar at Temple University and University of British Columbia. Currently, he is a full professor in the College of Computer and Information Technology at China Three Gorges University. He was a Lead Guest Editor of Pervasive and Mobile Computing, Special Session Chair for the 3rd International Conference on Internet of Vehicles (IOV 2016), and Local Arrangement Chair of I-SPAN 2018. He severs as Associate Editor for IEEE Access and EURASIP Journal on Wireless Communications and Networking. His research interests include VANETs, mobile social networks and opportunistic mobile networks.