

Can Your Friends Predict Where You Will Be?

Lei CAO, James SHE
HKUST-NIE Social Media Lab
Hong Kong University of Science & Technology
Email: {lcaoab, eejames}@ust.hk

Abstract—With the development of mobile device and wireless networks, user location becomes increasingly valuable in enhancing user experience, system performance and resource allocation. Location-based services have been not only an important perspective of social media, but also a significant contributor to big data analysis. Location prediction, as an interesting topic, can help improve system performance and user experience in location-based services. Existing algorithms on such prediction focus mostly on exploring regularity in users' movement history without taking advantage of the research on social networks, which can provide information on other factors such as peer influence in human mobility. In this work, the aim is to propose an enhanced location prediction model based on both users' mobility patterns and social network information and the proposed algorithm shows a significant improvement over existing ones.

Keywords—location prediction; social network analysis; big data

I. INTRODUCTION

Predicting the location of objects in mobile services has long been an intriguing research topic in data mining. Since most mobile devices are carried by men, the study on movement of devices is often replaced by the study on movement of human. With the surge of smartphones, location-related data collected from users have been enormous in volume and accumulating in a pace that it is difficult to apply traditional data analysis. The possibilities of extracting hidden information also add to the need for new data analysis strategies. The trend for location information analysis is merging into big data analysis.

A well description of user movement and an accurate prediction algorithm is critical in areas such as radio resources management, content delivery and location-based online services. Due to the limit on size and functionality of mobile devices, the digital messages sometimes can not be perceived perfectly. With the help of a precise movement prediction, the message can be delivered to a user not only through his personal devices, but enhanced by other devices in the surroundings where he's about to be. Public advertisements can also be delivered to its target group of users more precisely given a working prediction of those users' mobility.

Besides individual mobility, prediction of encounters between users is sometimes more important. One such case is contact-based forwarding systems like opportunistic routing

in *delay-tolerant networks* (DTN). In the Smart Doorplane Project [1], for example, a visitor can be notified about the probable next location of an absent office owner. The prediction is needed in this situation to decide how the visitor can meet the searched person quickly.

With recent development in social network analysis and social media, a user's physical activities can be better understood from his social network information. More insights could be given on how people's social life affect their mobility by associating their social network status with their physical status. On the other hand, a better knowledge of a user's movement in the physical world could help improve the functionality of online social networks. For example, a social network-based file sharing system could be implemented to work under flaky mobile ad-hoc network given that each user's mobility is known. Location prediction can also be used to enhance mobile content delivery and advertisement by precaching the needed content in target users' possible future locations.

A. Related Works

The architecture of DTN was first proposed in [2], which results a great interest in the research on prediction algorithms. A special case of DTNs is networks like Pocket Switched Networks (PSNs) [3], whose terminals are devices carried by people and dynamically networked. Hui et al. has made a research on the character of human mobility in terms of social structures and tried to use these structures to guide forwarding in PSNs [4]. In sociology and computer science, researchers have tried to model human mobility by *random walk* like Levy walk [5]. In addition to random mobility models, there are also extensive works in mobile motion prediction in the field of mobile communications. Mobile motion prediction algorithm [6][7] has been cited extensively in literature. In a lot of other applications, user's static location information is more beneficial than his mobility information. In [8], the model of multiple 2-D Gaussian distributions are used to model user's location. Researches on prediction of individual mobility are often related to resource management or data delivery. Peer influence and interaction in this situation is not an important factor. Encounter predictions, on the other hand, are studied mostly in contact-related communications [4][9][10]. The predictions are usually generated based only on their past

contact records. Users' individual behavior and life patterns are not considered. However, human mobility should be the mixture of influences from both his own daily life patterns and social interactions. In [11], a location prediction scheme based on social influence is proposed. With the help of social correlation, each user is coupled with a Most Possible Companion (MPC) at a particular time based on his past records and social relationship. Then the user's location is predicted based on his MPC's location. This method is referred to as *MPC Method* in later document.

In this work, a human mobility model to predict user's future locations is proposed. The contribution of this work can be summarized as follows:

- A novel model for depicting and predicting user location is proposed which incorporates both his regular mobility patterns and social influence;
- Unlike existing methods using social information for location prediction, the closeness between friends is considered to grant different weights on different friends based on interaction frequencies in the proposed method;
- While many other methods work with the continuous location data like GPS coordinates, the proposed model works with discrete data like cell tower ID and AP ID with a variant version adapting to continuous data;

The rest of the paper will be organized as follows. In Section II, the mathematical model of the proposed prediction algorithm is described. Simulation results based on the given algorithm are presented in Section III.

II. PROPOSED LOCATION PREDICTION MODEL

There are two major notations for location information, one is global positioning system (GPS) coordinates and the other is locations described by label IDs such as cell tower or wireless access point (AP). The two location labels carry different information and are usually dealt with using different approaches. The coordinates data are continuous and location information carried is more precise. To handle these data, on the other hand, requires algorithms dealing with probability densities. The location prediction algorithm proposed in [8] is dealing with such data. The cell-id-labelled data are discrete and the precision of the recorded location depends on the radius of the cell coverage. In this work, the focus is on the latter notation for two reasons: it is easier to interpret as logical location than coordinates and it is more applicable as most of mobile device access Internet via grid-structured networks.

Although human behaviors always show a strong regularity on time and places, they are affected by other external factors. In the past, people have showed that social relationships have huge impacts on human mobilities and it is a reliable element for prediction since relationships are fairly stable. Researchers have utilized social relationships and interactions for certain predictions and decisions such as

forwarding in DTNs. In [5], authors have shown that human mobility is highly correlated to social life and it is a valid assumption that location prediction can be made based on one's social relationships. In the proposed algorithm, the Periodic and Social Location Prediction Model (PSLPM), these two factors influencing user location are represented by a combination of a non-social part, Periodic Location Prediction Model (PLPM) and a social part, Social Location Prediction Model (SLPM) respectively.

A. The Periodic Part (PLPM)

Consider a scenario where a mobile user is connected to a cell tower and his current location is labelled with the cell id. The user may move away from this cell and enters an adjacent cell after a brief interval of time. The fundamentals of a working prediction scheme is to predict in which cell a person is most likely to be and which cell he will enter after a period of time. The nature of this problem and the hierarchical character of network structure lends itself to adopting Markov Chain model [12]. Markov chain model has been adopted in wireless networks in the past to model transitions between wireless APs [13]. To the best of our knowledge, this paper is the first one investigating using Markov chain model for user location prediction incorporated with social network information.

The topology of cellular system can be abstracted into a graph similar to standard Markov chain in Fig. 1(a). In this setting, the state automatically corresponds to the cell and the transition probability between states represents the probability of a user migrating from one cell to another. The user location history can thus be noted by a sequence of state transitions. With the location history, the transition probability between cells can be calculated and a complete Markov chain can be established.

For a given user, assume L is the set of all his friends' location records. Let L_n and t_n denote the last known location and time of n -th friend and w_n is his share of influence on the movement of the user. All time parameters are denoted as t . Despite the continuous nature of t , discrete values have to be taken in calculations. The temporal resolution of t is 1 minute.

In this paper, a Markov chain is proposed as follows. $S = \{s_1, s_2, s_3, \dots, s_M\}$ are the M states or cells in the system $\Pi_t = \pi_i$ is the stationary state probabilities. π_i denotes the stationary probability of state i $A = \lambda_{ij}$ are the state transition probabilities where λ_{ij} indicates the probability of moving from state i to j .

$$\lambda_{ij} = P(t_k = s_j | t_{k-1} = s_i) \quad (1)$$

1) *Learning Process*: Unlike standard Markov chain models, transition probabilities might change periodically in human mobility as the probability of a person migrating from one location to another might be different as time

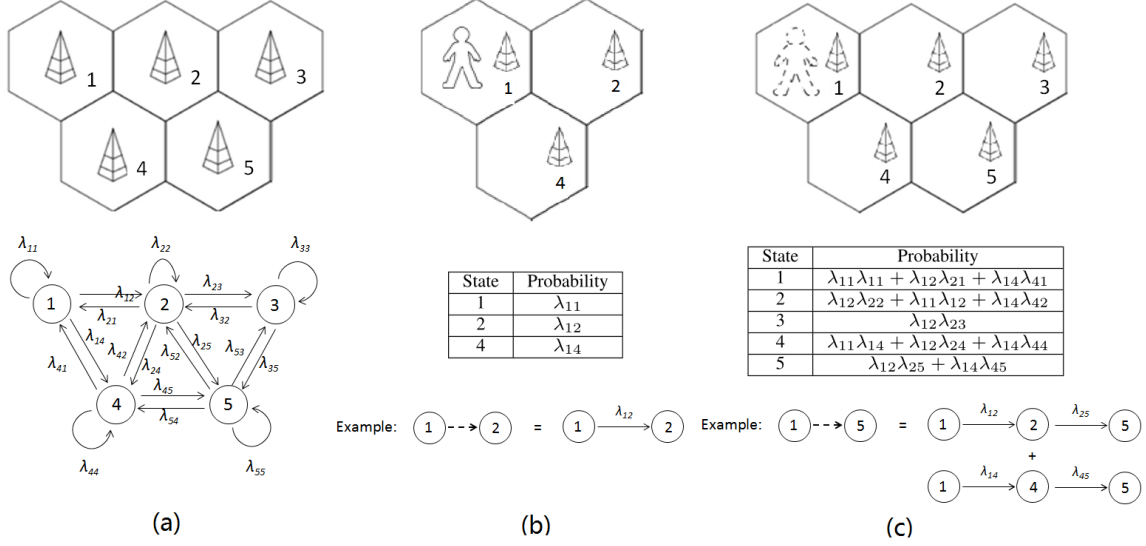


Figure 1. Markov chain model: (a) Cell network state transition; (b) Given previous record 1 timeslot ago; (c) Given previous record 2 timeslots ago

changes. Therefore dynamic transition probabilities are required in this model. A_t denotes the transition probability matrix at time t . The resolution of t determines on a large scale how accurate and specific this model can predict. In this work, the transition probabilities are sampled every hour, or $A_t, t \in \{0, 1, 2, \dots, 23\}$.

2) *Prediction Process*: The prediction with Markov chain can be categorized into two parts. Without knowledge of any recent appearance of the user, or given that the last record of the user is fairly long ago, stationary probability can be used as the probability of user being at corresponding location.

$$P(x_t = s_i) = \pi_{i,t} \quad (2)$$

The other case is when a recent record of the user is available. If the most recent record shows that the user is at s_i at t_0 , the probability of user being at s_j at t can be updated as

$$P(x_t = s_j | x_{t_0} = s_i) = \{B\}_{ij} \quad (3)$$

where $B = A^{t-t_0}$

B. The Social Part (SLPM)

Apart from periodic patterns stated in PLPM, it is safe to assume that part of one's location records follow those of his friends' as a result of social influence. Given a precisely defined social graph, location prediction could be enhanced by one's friends' location information. This possible enhancement is facilitated by appending a friend-influenced component to the original PLPM.

For a given user, let $\sum w_n = 1$ and probability of user being at s_j at time t can be expressed as

$$P(x_t = s_j | x_{t_0} = s_i, L) = w_0 \{B_0\}_{ij} + \sum_L w_n \{B_n\}_{L_n j} \quad (4)$$

where $B_0 = A^{t-t_0}$ and $B_n = A^{t-t_n}$.

The above model is a combination of both non-social and social part of the model. Although intuitively the weight in above equation depends solely on the closeness, or social tie as defined in some works, between users, it is difficult to retrieve accurate values of w_n in reality. Instead, social correlation is used to represent closeness between users and this model is simplified by only taking the most recent record of a most influential friend. The social correlation between two users u and v are defined as

$$C_{u,v} = \frac{\sum_t I_{u,v}(t)}{\sum_t I_{u,u}(t)} \quad (5)$$

where

$$I_{u,v}(t) = \begin{cases} 1, & u \text{ and } v \text{ are at the same location at time } t \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

It is important to note that the social correlation is not necessarily an exact measurement of how close friends are, but it is valid to indicate the closeness and resemblance between friends in terms of their trajectories. Although accuracy is compromised in this approach, computational efficiency is hugely improved. Assume $y_{t'}$ denotes the record of that friend in F , which is taken at time t' . The complete PSLPM is as below:

$$P(x_t = s_j | x_{t_0} = s_i, y_{t'} = s_k) = (1-pC') \{B\}_{ij} + pC' \{B_n\}_{kj} \quad (7)$$

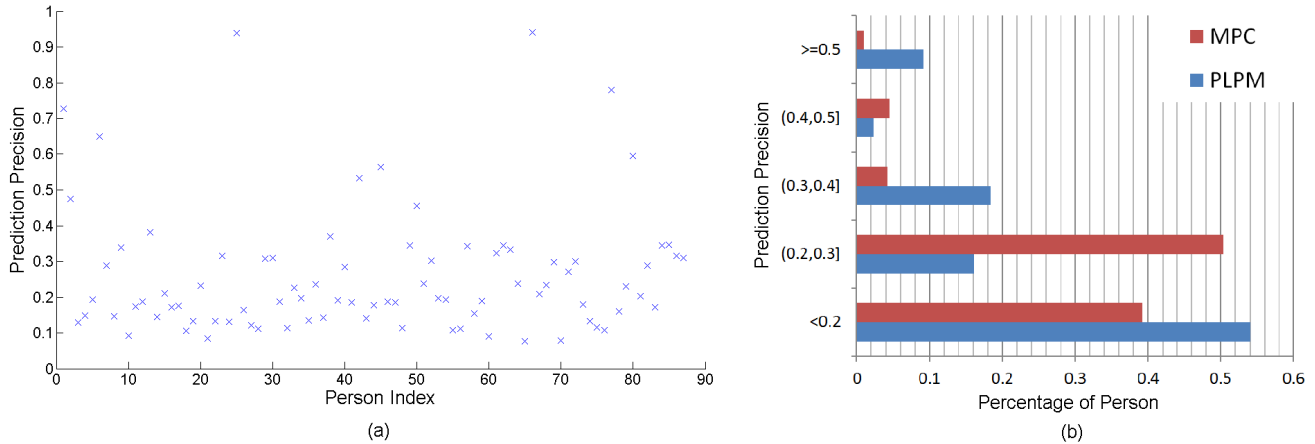


Figure 2. (a) Prediction precision of PLPM; (b) PLPM in comparison with MPC

where $B = A^{t-t_0}$, $B_n = A^{t-t'}$ and $p = \frac{t'-t_0}{t-t_0}$ ($t \geq t_0$).

t' and $y_{t'}$ are retrieved by the method shown in the pseudocode below.

The value of pC' represents how much the user's mobility and location is influenced by his friends and it would largely affect the performance of the model. Two factors are used to determine the value of pC' : C' , the resemblance of trajectory between the user and his closest friend, and p , which is a parameter evaluating how valuable this friend's location record is in comparison to the time of user's own last known location. p should decay as time goes by, and in this work, it is assumed that it's decaying linearly and represented by $\frac{t'-t_0}{t-t_0}$.

Algorithm 1 Simplified SLPM

```

1: while  $L_n \in L$  do
2:   if  $t_n > t_0$  and  $C_n > C'$  then
3:      $C' \leftarrow C_n$            ▷ identify the closest friend
4:      $t' \leftarrow t_n$ 
5:      $y_{t'} \leftarrow L_n(t_n)$   ▷ replace friend's most recent
    record
6:   end if
7: end while
8: return  $t', y_{t'}, C'$ 

```

It is also noteworthy that when the location of a user's friends are unknown or when the user's own last known location is more recent than any of his friends', PSLPM shrinks to PLPM as in (3) where the social part of the model becomes useless. On the other hand, when the location history of the user himself is unknown ($t - t_0 \rightarrow \infty$), or the current location of the friend is known ($t' = t$), PSLPM can also be simplified where $pC' = C'$.

III. EVALUATION RESULTS

The proposed prediction scheme is tested on a real cellular network trace. Details about the data set, data processing, prediction and result analysis will be discussed in this section.

A. Dataset and Metrics

The dataset used in this simulation is from Reality Mining Project [14] of MIT Media Lab. 106 cellphones are involved in this project and numerous data about each phone and related individual are collected. Although this dataset provides many aspects of individual's social life through mobile network, the data needed in this work is the location records of each user and relationship between users. Location history in this project is recorded as sequences of cell IDs the user traversed with Coordinated Universal Time (UTC) timestamps in resolution of 1 second. Of all 106 data entries, 87 are valid for the simulation.

Prediction Precision (P) is used to evaluate the performance of this model, which is given as follows:

$$P_i = \frac{\text{number of correct prediction}}{\text{number of total prediction}} \quad (8)$$

B. Periodic Location Prediction

In the first simulation set, PLPM with selection of users in the dataset is used for evaluation. In comparison, the prediction precision of MPC method in [11] is also used. For each user, the data is divided into training data and testing data. After feeding training data to extract transition matrix, other samples are used in testing data to evaluate prediction precision. The first experiment evaluates the prediction precision of all participants without prior knowledge. In other words, the stationary probabilities are used for prediction for all users. The results are shown in Fig. 2 and majority of users get a prediction precision around 20%. In comparison, the prediction is slightly better than the results from MPC, in

terms of average prediction precision, which is 25%. PLPM is also able to provide better predictions with users who have more regular patterns as it performs much better in the high precision range.

In the second experiment, the performance of the model when prior locations of a person are provided is examined. After training the model with training data of a user, an excess recent location is also available which is recorded Δt before prediction in terms of *minutes*. The range of Δt is from 5 to 30 minutes and speculation is that with the more recent location record, the higher precision can be achieved. The statistical results for prediction precision of users's data is shown in Fig. 3 together with SLPM.

C. Recency of Friends in SLPM

In the second simulation set, the focus is to evaluate the performance of the PSLPM. As explained in Section III, the most recent location record of the most influential friend in terms of social correlation is used to enhance prediction. In the dataset, however, the social graph is quite scarce and social correlation is in general not very high for any given user. In this simulation, the time of the selected user's friends' last known record is labelled as t' and the time difference between that and prediction is $\Delta t' = t - t'$. The results are shown in Fig. 3. Although the improvement is not as significant as expected, mostly due to the limited size of the social network, it is observed that a social influence taken into consideration for location prediction indeed help enhance the prediction results.

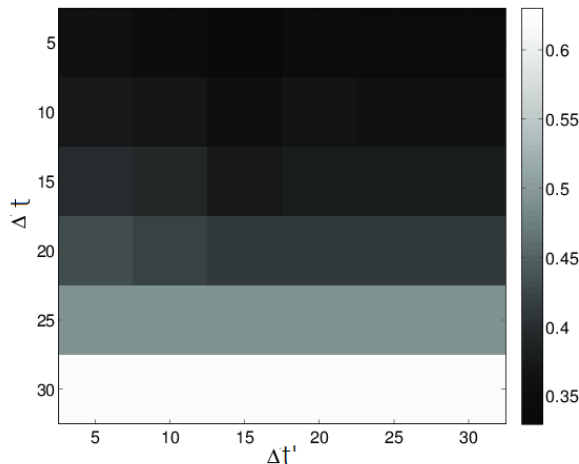


Figure 3. Prediction precision of PSLPM

D. Granularity and Prediction Accuracy

As mentioned in previous paragraph, the radius of the cell would have a large influence on the prediction precision of the PSLPM. With data from large-ranged cells, an intuitive guess would be the data yield a better prediction precision

than those from small-range cells. In the dataset used in this simulation, the cell id are composed of two segments: the first segment is area ID and the second is cell ID. The area IDs can be considered as a cell with larger radius. These data of area IDs are used in comparison to cell IDs to prove the validity of previous hypothesis. The prediction results of both calibers are shown in Fig. 4. The prediction precision with area IDs are much better than that with cell IDs, resulting in conclusion that the cell with larger radius yields better prediction precision. However, the larger cell covers more area, which provides less detail of the exact location of the user. Hence the prediction precision can be improved with larger cells in expense of granularity. In real life, locations are usually depicted by labels such as "shopping center", "station", "school" and etc. When locations of smaller size, or the more specific labels are used, the difficulty for prediction will rise as precision will drop according to the results. For instance, prediction of whether a person is on campus tend to yield less prediction error than that of whether he is in a specific classroom. However, the successful prediction of the latter occasion is usually more informative.

IV. CONCLUSIONS

In this paper, a location prediction model based on both persons periodic movements and social relationships is proposed. This model is based on the understanding that a person's mobility is under mixed influence of his life patterns and occasionally peer interference. It also works perfectly with the common location information of cellular networks. Experiments on a real data trace showed a good performance and a significant improvement over existing algorithms.

In the future work, improvement of predicting precision could be potentially achieved if weighted relationships are taken into consideration. At the same time, algorithms of better and higher performance for information mining, location and person prediction and strategy for the use of these algorithms would be developed. Meanwhile, an implementation of a location prediction system utilizing the same method is under development on smartphones. In the next work, the performance of this algorithm in real environment will be tested with real-time information retrieval and analysis.

REFERENCES

- [1] W. Trumler, F. Bagci, J. Petzold, and T. Ungerer, Smart doorplate, *Personal Ubiquitous Comput.*, vol. 7, no. 3-4, pp. 221-226, Jul. 2003. [Online]. Available: <http://dx.doi.org/10.1007/s00779-003-0218-y>
- [2] K. Fall, A delay-tolerant network architecture for challenged internets, in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, 2003, pp. 27-34.

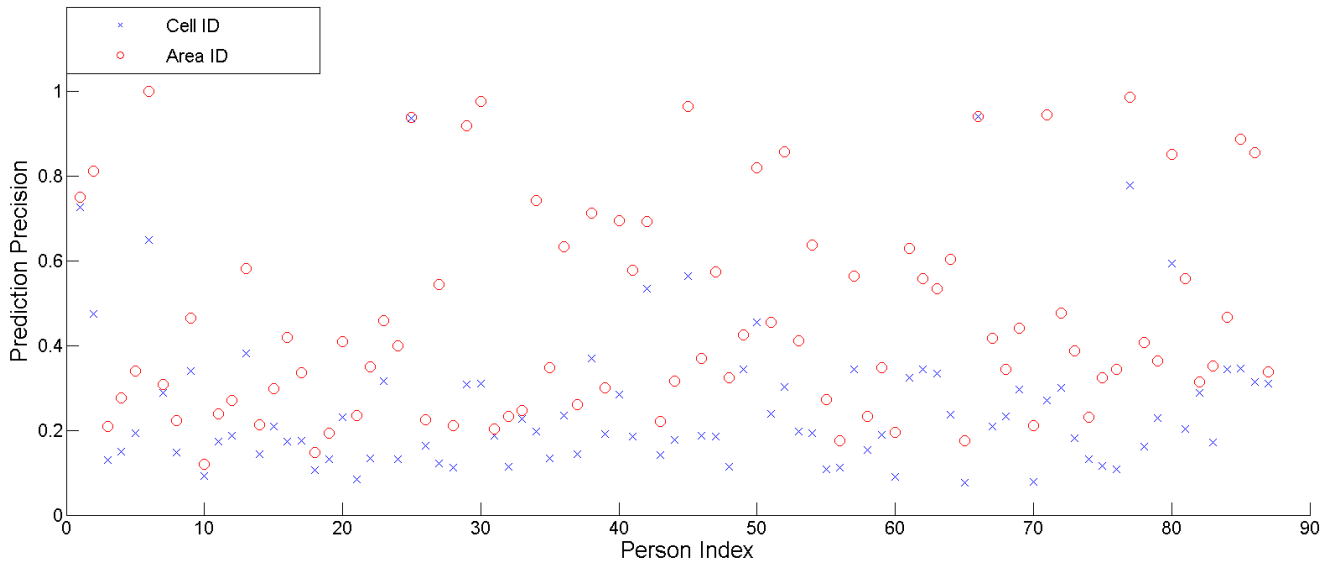


Figure 4. Prediction precision enhancement of using Area ID instead of Cell ID

- [3] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, Pocket switched networks and human mobility in conference environments, in *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, 2005, pp. 244-251.
- [4] P. Hui, J. Crowcroft, and E. Yoneki, Bubble rap: social-based forwarding in delay tolerant networks, in *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, 2008, pp. 241-250.
- [5] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, On the levywalk nature of human mobility, *Networking, IEEE/ACM Transactions on*, vol. 19, no. 3, pp. 630-643, June 2011.
- [6] G. Liu and G. Maguire, A class of mobile motion prediction algorithms for wireless mobile computing and communication, in *Mob. Netw. Appl.*, 1996, pp. 113-121.
- [7] T. Liu, P. Bahl, and I. Chlamtac, Mobility modeling, location tracking, and trajectory prediction in wireless atm networks, in *IEEE Journal on Selected Areas in Communications*, vol. 16, 1998, pp. 922-936.
- [8] E. Cho, S. A. Myers, and J. Leskovec, Friendship and mobility: user movement in location-based social networks, in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD 11. New York, NY, USA: ACM, 2011, pp. 1082-1090.
- [9] T. Hossmann, F. Legendre, and T. Spyropoulos, From contacts to graphs: Pitfalls in using complex network analysis for DTN routing, in *INFOCOM Workshops 2009, IEEE*, 2009, pp. 1-6.
- [10] D. Fischer, D. Basin, and T. Engel, Topology dynamics and routing for predictable mobile networks, in *IEEE International Conference on Network Protocols*, 2008, pp. 207-217.
- [11] Y. Gong, Y. Li, D. Jin, L. Su, and L. Zeng, A location prediction scheme based on social correlation. in *VTC Spring*. IEEE, 2011, pp. 1-5.
- [12] H. SI, Y. WANG, J. YUAN and X. SHAN, Mobility Prediction in Cellular Network Using Hidden Markov Model, in *7th IEEE Consumer Communications and Networking Conference (CCNC)*, 2010
- [13] P. S. Prasad and P. Agrawal, Movement prediction in wireless networks using mobility traces, in *Proceedings of the 7th IEEE conference on Consumer communications and networking conference*, ser. CCNC10. Piscataway, NJ, USA: IEEE Press, 2010, pp. 714-718.
- [14] N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure using Mobile Phone Data", in *Proceedings of the National Academy of Sciences*, 106(36), pp. 15274-15278. 2009