RESEARCH ARTICLE OPEN ACCESS

Centrality Prediction in Mobile Social Networks

Poria Pirozmand*, Guowei Wu, Xinjiao Li
School of Software, Dalian University of Technology, Dalian 116620, China

Abstract
By analyzing evolving centrality roles using time dependent graphs, researchers may predict future centrality values. This may prove invaluable in designing efficient routing and energy saving strategies and have profound implications on evolving social behavior in dynamic social networks. In this paper, we propose a new method to predict centrality values of nodes in a dynamic environment. The proposed method is based on calculating the correlation between current and past measure of centrality for each corresponding node, which is used to form a composite vector to represent the given state of centralities. The performance of the proposed method is evaluated through simulated predictions on data sets from real mobile networks. Results indicate significantly low prediction error rate occurs, with a suitable implementation of the proposed method.
Keywords: dynamic networks, centrality prediction, social networks, mobile networks

I. Introduction

Networks undoubtedly play an important role in the modern society. Perhaps the most noteworthy reason of accelerated progress in science and technology is the increased growth, availability and reliability of information networks. Information networks, not only includes the *World Wide Web*, but also other networks such as scientific collaborations, communication networks, among numerous other forms. The concept of a' social network' may be as old as human civilization itself; however, it has not been extensively investigated until recent years, due to advancements in the communication technology, international trading, and collaborative research projects.

The mathematical analysis of social networks has a fundamental basis in Graph Theory, which makes it possible to model, analyze and illustrate nodes and relationships effectively. The methods developed to analyze social networks have been presented. Several general problems in the context of social and mobile networks have been investigated include community detection, centrality measures, network complexity, key player identification, information routing strategies, epidemic flows among others[1].

In this paper we analyze the evolution of centrality in different nodes and propose new methods for the prediction of future centrality values. The performances of which shall be evaluated using several real opportunistic mobile network datasets.

This paper is organized as follows: Related work concerning dynamic networks and centrality in social networks is discussed in section 2. In section 3 the main concepts of interest of this work are presented which include graph representation of networks in subsection 3-1, dynamic networks in subsection 3-2, and the centrality measures implemented in our study in subsection 3-3. In section 4 the proposed centrality prediction method is discussed.Section5 on three real datasets and by calculating the prediction errors; the performance of the proposed algorithm is shown. Concluding remarks are presented in section 6.

II. Related Work

Identifying node centrality in networks with greater accuracy may have unprecedented benefits, as it may allow one to design better routing strategies or to select key players in a more appropriate way. It is also useful in the context of identifying criminal and terrorist networks [2].Various metrics and algorithms have been proposed to quantify node centrality in social and mobile networks. However, most of these method shave been formulated to study static networks in which the nodes and links do not change with time. It is obvious that dynamic networks pose significantly more challenges. Such changes in dynamic networks may occur due to link removal or link appearance between nodes [3-7], or due to the mobility of the nodes themselves [8-10]. The study of dynamic social networks [7] and mobility analysis in mobile networks [11, 12] has attracted increasing interest by researchers in recent years. Particularly, opportunistic delay tolerant networks (DTNs) have been extensively investigated in recent years by researchers in telecommunication and computer engineering [13-19].The use of specific social aspects in networks, including node centrality, for the information routing in DTNs has been a rich and promising research direction [17].

In a dynamic network, it is very common that the future level of centrality of nodes would not be the same as the current. By postulating on the behavior of the network and predicting possible future nodes of centrality, one may establish a more efficient routing or energy consumption strategy for the network. This of course is applicable to both mobile communication and sensor networks. The problem of centrality prediction in dynamic social or mobile networks has been recently addressed in several works [1,20]. In [1], several simple methods for centrality prediction are compared and the effects of various parameters on the performances are investigated. An iterative matrix model for dynamic networks, which particularly analyses the evolution of node centrality, has been proposed in [20].

III. Centrality in Dynamic Network

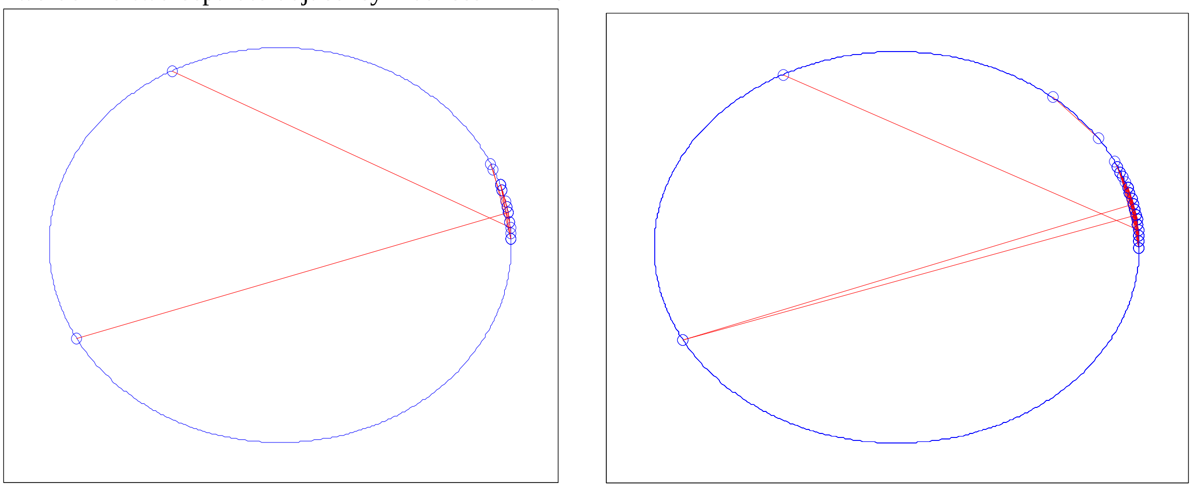
In this section the general formulation of the problem is described. The notions of dynamic network in the context of graph theory, evolution of the network, and centrality are addressed.

3-1- Network Graph

A static graph representing a network of a set V of nodes $v_i \in V$and a set E of links $e_{i,j} \equiv (v_i, v_j) \in E$between them is shown by $G = (V, E)$. The links in E can be directed which means$(v_i, v_j) \neq (v_j, v_i)$. In most applications implemented on contact networks, such as those used in this study, the network is modeled by directed graphs where a link represents a contact between a 'sender' and a 'receiver', or an observing device and an observed device. The adjacency matrix A of the directed graph is defined as a$|V| \times |V|$ matrix in which the element $A[i,j]$ is zero if$(v_i, v_j) \notin E$, 1 if $(v_i, v_j) \in E$ and is -1 if$(v_j, v_i) \in E$. Note that a restriction of only one link was defined to exist between two nodes at a given time frame. The case of multilink contacts can be simply handled if we define two separate adjacency matrices A^+ and A^- for sending and receiving links respectively. Then the number of sending links from v_i to v_j is set,$A^+[i,j]$ and the number of receiving links to v_i from v_j is set, $A^-[i,j]$.

3-2- Dynamic Network

In dynamic networks, we can observe the network as a sequence of snapshots of static graphs in discrete time intervals. Therefore, we can have several time dependent adjacency matrices like, A_t^+. In this manner,$A_t^+[i,j] = 1$ means that there is a link from v_i to v_j at time instance t. In practice, this definition may be not very useful in some networks with a high frequency of changes, such as mobile communication networks. As such, this approach is useful for networks with rather slow changes in configuration such as friendship networks. The reason is that in a rapidly changing network, there may be several link appearances or removals for a given time window $w_t = [t, t + w]$ which should be noticed in an aggregated way in the analysis of temporal centrality. For simplicity, consider a snapshot of the network from Cambridge dataset [23] as shown in figure 1a. At roughly the same time interval, the aggregated network in a 20 second time window is shown in figure 1b. As it is seen, several contacts have appeared in a short period of 20 seconds. For centrality analysis and prediction, this time window consideration is very important, so we utilize windowed adjacency matrices $A_{w_t}^+$and $A_{w_t}^-$ in our study.



(a) (b)

Fig.1: Difference between (a) a single snapshot of network, and (b) aggregated graph of network in a 20 sec time window around the same time, from Cambridge dataset

3-3- Centrality Measure

In this paper we use Katz centrality measure [20] which sums weighted paths between one node and another, such that longer paths have lower weights while the most direct paths (links) possess a higher weight. From the adjacency matrix A, the Katz vector of centrality measure can be calculated as:

$$C = \left(\sum_{k=1}^{\infty} \sum_{j=1}^{|V|} (\alpha A)^k\right) \mathbf{1}_{|V|} \qquad (1)$$

where $\mathbf{1}_{|V|}$ is a vector of length $|V|$ consisting of 1's, and α is a small weight factor. By selecting a low enough value for α (smaller than the inverse of largest absolute Eigen value of the adjacency matrix), the summation may be convergent. For our study, we choose $\alpha = \frac{1}{|E|}$ which gives a centrality measure of 1 for a node in the center of a star graph. We further calculate the summation up until $k = 5$ and suppress further terms.

We also consider the degree of nodes in our centrality analysis. The degree can be seen as the first term in Katz centrality ,if $\alpha = 1$. By using the degree of nodes we are able to differentiate between center nodes of star graphs with a varied number of links. One of the advantages of using these measures (i.e. Katz centrality and degree centrality) is the speed of computation in contrast to other measures founded on shortest path detection.

IV. Centrality Prediction

To predict future centrality levels of nodes in a network, we utilize correlation and distance analysis on previous centrality measurements. The index we use for centrality is the product of Katz centrality values and degree centrality values. The reason of this choice is that the Katz centrality alone (calculated by setting $\alpha = \frac{1}{|E|}$), has been proved inefficient in calculating centrality in star graphs of varied degrees. On the other hand, degree centrality alone does not contain information regarding non-direct paths between nodes. Therefore, we define a centrality metric such as:

$$\hat{C}_i = C_i \deg(v_i) \qquad (2)$$

It is proposed that future values of this measure can be predicted from previous values. Furthermore, in contrast to previous methods, the proposed approach utilizes the previous values of all nodes for an accurate prediction of centrality, not merely the history of the node in question.

The proposed algorithm for prediction of future centrality values can be described as:

1- At each time t, calculate centrality measures $\hat{C}_i$ for all nodes and store as vector $\hat{C}(t)$.

2- At time t_n , calculate correlation coefficients between $\hat{C}(t_n)$ and all the saved $\hat{C}(t_i)$ vectors for $t_i = t_n - l_i$, where $l_i = t_n - t_i$ is the lag. Save correlation values in an array $corr(i)$. Also, calculate Euclidean distance between $\hat{C}(t_n)$ and every saved $\hat{C}(t_i)$ vectorand store in array $d(i)$.

3- Find the smallest index i_c for which $corr(i_c) = \max\{corr\}$. Also find smallest index i_d for which $d(i_d) = \min\{d\}$.

4- Set the centrality vector $\hat{C}(t_{i_c+1})$ as the predicted centrality vector $\hat{C}_c(t_{n+1})$. Set the centrality vector $\hat{C}(t_{i_d+1})$ as the predicted centrality vector $\hat{C}_d(t_{n+1})$.

5-After the occurrence of $\hat{C}(t_{n+1})$, calculate the error between real values and predicted values.

As it can be seen, this algorithm calculates correlation and distance between vectors composed of centrality values for every node. After which the state of the whole network is compared to its previous state. So the prediction of centrality measure for a node is obtained from the historical information of all nodes. To provide the definition of distance and correlation used, consider the array $\hat{C}(t_i)$ formed by placing $\hat{C}_j(t_i)$ values as its elements, such that:

$$d(i) = \sqrt{\sum_{j=1}^{|V|} (\hat{C}_j(t_n) - \hat{C}_j(t_i))^2} \qquad (3)$$

and

$$corr(i) = \frac{\sum_{j=1}^{|V|} \left(\hat{C}_j(t_n) - \overline{\hat{C}(t_n)}\right)\left(\hat{C}_j(t_i) - \overline{\hat{C}(t_i)}\right)}{\sqrt{\sum_{j=1}^{|V|} \left(\hat{C}_j(t_n) - \overline{\hat{C}(t_n)}\right)^2} \sqrt{\sum_{j=1}^{|V|} \left(\hat{C}_j(t_i) - \overline{\hat{C}(t}\right.}} \qquad (4)$$

where the "bar" on a quantity means its mean value.

V. 5- Performance Evaluation

To show the performance of the proposed centrality prediction method, we deploy it to three real datasets, namely; Haggle project named Intel, Cambridge, and Infocom [21-23]. The Intel dataset consists of contact recordings of 9 I Mote devices for a period of 5 days in the Intel Research Cambridge Corporate Laboratory. Cambridge dataset contains contact recordings of 12 I Mote devices for 5 days in the Computer Lab at University of Cambridge. Infocom dataset contains the contacts of 41 devices for 3 days during the IEEE Infocom conference at The Grand Hyatt in Miami.

To show the capability of correlation and distance analysis in the proposed prediction method, the plot of correlation between $\hat{C}(t_n)$ and $\hat{C}(t_i)$ vectors for $t_i = t_n - l_i$, versus the lag l_i is depicted in figure 2 using a case from Infocom dataset with an aggregate time window size of 120. For the same case, the plot of distance between $\hat{C}(t_n)$ and $\hat{C}(t_i)$ vectors is shown in figure 3.

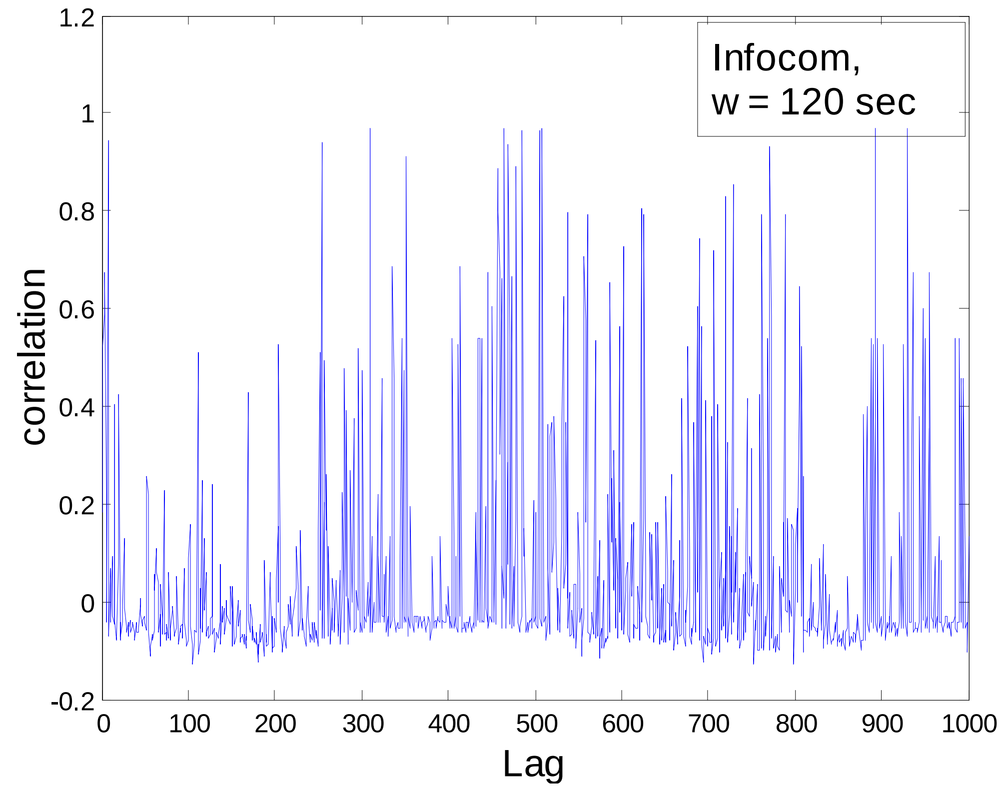


Fig.2: Correlations between current and past vectors of centrality values.

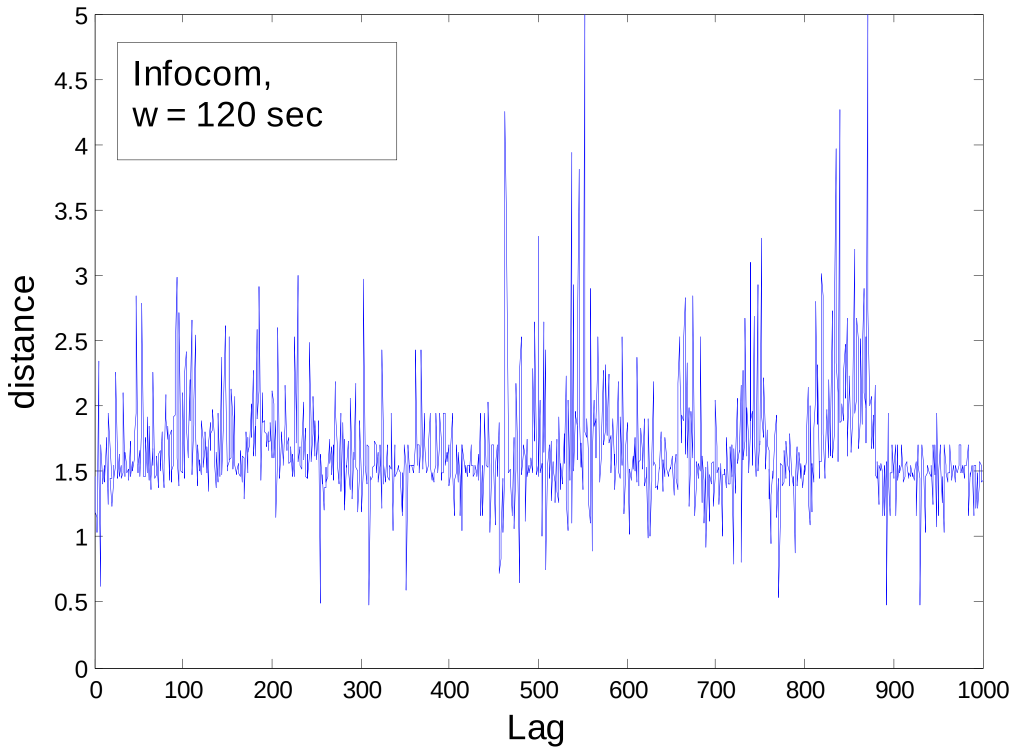


Fig.3: Distance between current and past vectors of centrality values

In figure 2, a high correlation of values is observed for several lags. Figure 3 highlights the existence of several low distances between current and past centrality vectors. Thus, substantiating the premise of the proposed algorithm. It should be noted that the result is shown for a t_n, with a suitably nonzero neighborhood of centrality values. This is important because for zero centrality periods, the correlation and distance are not meaningful in this context.

The results of deploying the proposed prediction method on three datasets, using various aggregation time windows are shown in Table 1 where e_c represents mean squared error (MSE) for prediction of observed correlations, and e_d is used to indicate MSE for prediction of distance. MSE is defined as

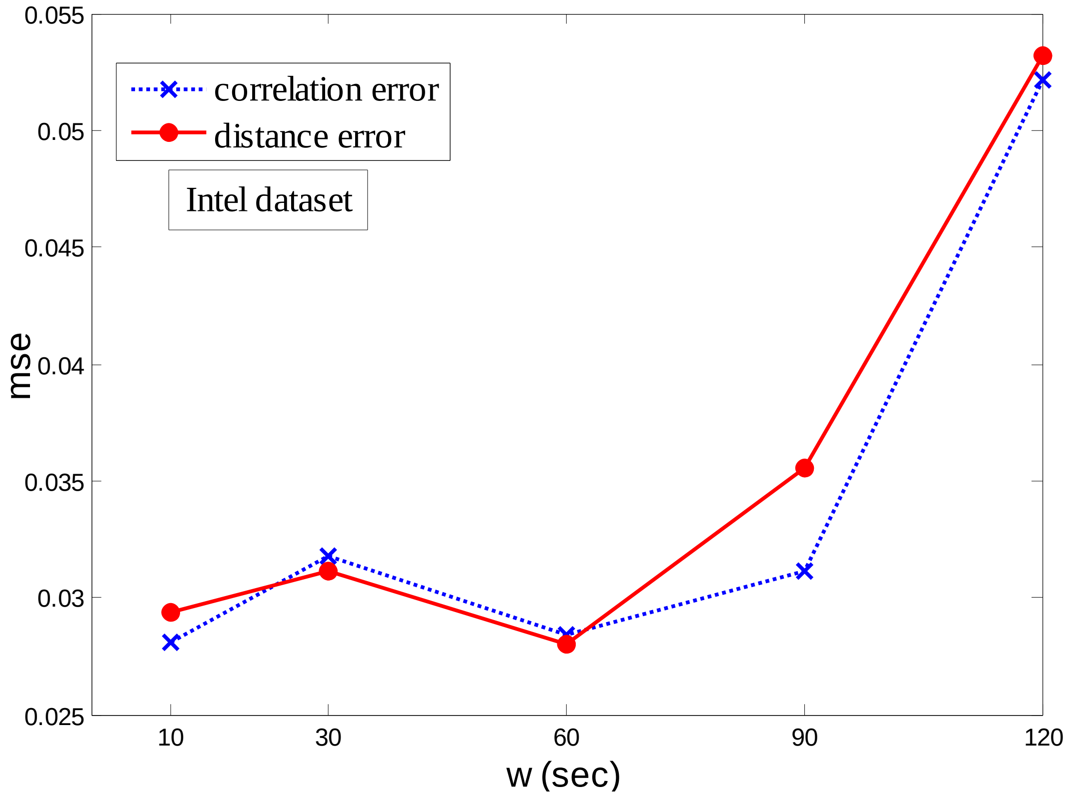
$$MSE = \frac{\sum_{j=1}^{|V|}(\hat{C}_j(t_{n+1}) - \tilde{\hat{C}}_j(t_{n+1}))^2}{|V|} \quad (5)$$

where $\tilde{\hat{C}}_j(t_{n+1})$ is the predicted value.

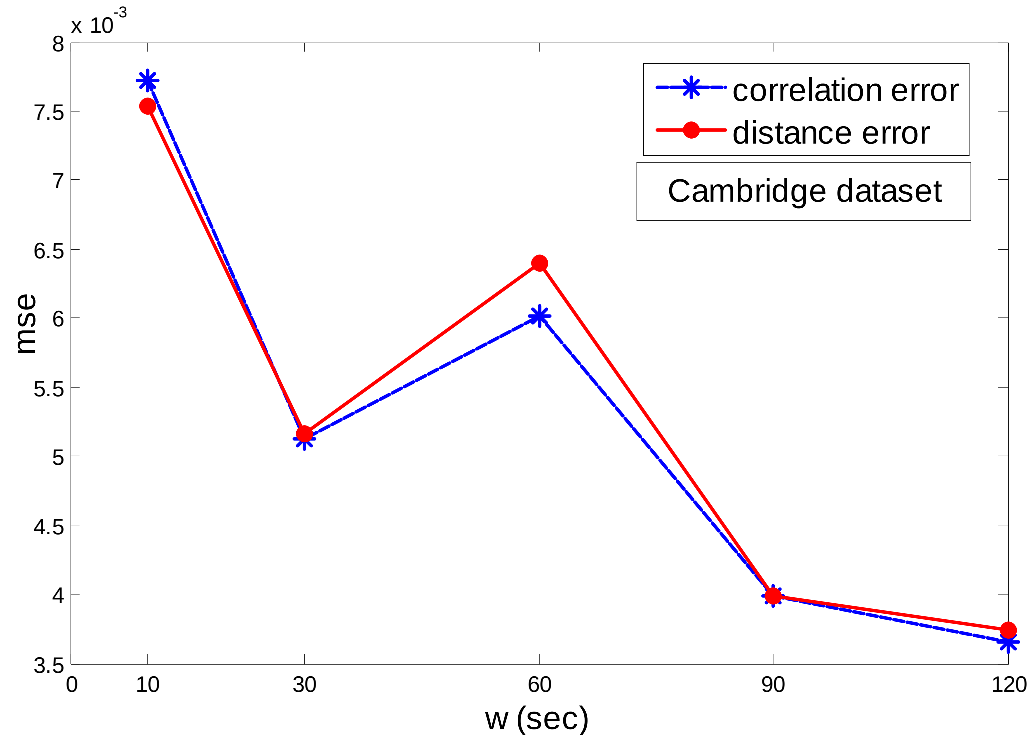
It can be observed that for each dataset certain window sizes indicate better results through correlation and others indicate distance analysis yield better results. However, these differences between the results from two kinds of prediction methods are almost negligible. It is also seen that for each dataset a suitable aggregation window size can be set, which is mainly dependent to the nature of the experiment situations behind dataset itself. Variations of correlation and distance mean squared errors by changing the window size are shown in figure 4.

case \ *w* (sec)		10	30	60	90	120
Intel	e_c	0.028085	0.031759	0.028411	0.031098	0.052171
	e_d	0.029357	0.031117	0.028026	0.035527	0.05317
Cambridge	e_c	0.0077186	0.0051221	0.0060191	0.0039887	0.0036592
	e_d	0.0075392	0.0051669	0.0064027	0.0039887	0.0037444
Infocom	e_c	0.015124	0.012267	0.0073041	0.0075916	0.0087773
	e_d	0.013737	0.013783	0.0078664	0.0066497	0.0078577

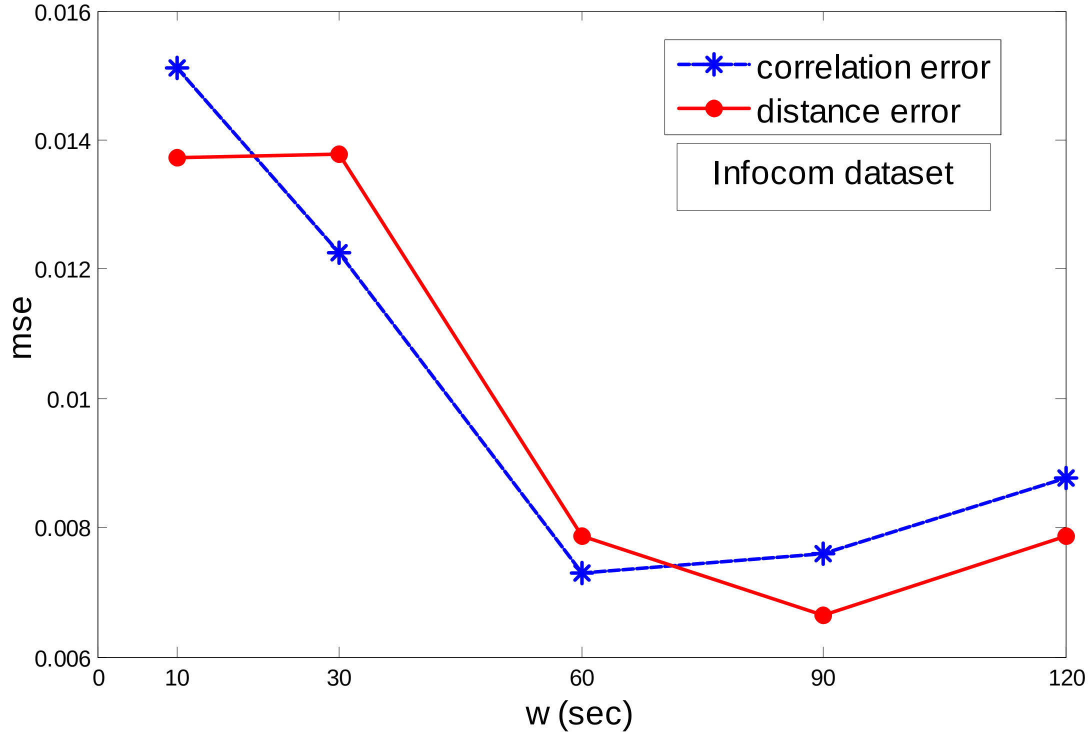
Table1. Results of prediction errors by correlation and distance based analysis for the three datasets and different aggregation time window sizes.



(a)



(b)



(c)

Fig.4: Variation of correlation and distance MSE by window size for (a) Intel, (b) Cambridge, and (c) Infocom datasets.

VI. Conclusion

This paper has proposed a prediction method for centrality of nodes in dynamic social mobile networks. This method is based on observing the correlation and distance values between the current vector of node centrality and its past measurements. As such, a given state of the network can be compared to its previous states. It has been found that correlation values are high and distance values are low for similar states of the network. The future node centralities are predicted by finding the nearest suitable lag, based on maximum correlation and minimum distance. In this study, the node centralities are quantified as a product of Katz centrality and the simple degree of the nodes. The performance of the proposed method is evaluated by performing predictions on three datasets of real mobile social networks. By both correlation and distance-based prediction methods, very low prediction errors are obtained for the three datasets, with various aggregate time window sizes.

References

[1] H. Kim, J. Tang, R. Anderson, and C. Mascolo, "*Centrality prediction in dynamic human contact networks*," *Computer Networks*, vol. 56, no. 3, pp. 983–996, 2012.

[2] X. Liu, E. Patacchini, Y. Zenou, and L.-F. Lee, "*Criminal networks: Who is the key player?*," 2012.

[3] B. Skyrms and R. Pemantle, "*A dynamic model of social network formation*," in *Adaptive Networks*, Springer, 2009, pp. 231–251.

[4] D. Greene, D. Doyle, and P. Cunningham, "*Tracking the evolution of communities in dynamic social networks*," in *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*, 2010, pp. 176–183.

[5] N. P. Nguyen, T. N. Dinh, S. Tokala, and M. T. Thai, "*Overlapping communities in dynamic networks: their detection and mobile applications*," in *Proceedings of the 17th annual international conference on Mobile computing and networking*, 2011, pp. 85–96.

[6] A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro, "*Time-varying graphs and dynamic networks*," in *Ad-hoc, Mobile, and Wireless Networks*, Springer, 2011, pp. 346–359.

[7] J. L. Pfaltz, "*A mathematical model of dynamic social networks*," *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 863–872, 2013.

[8] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang, *"A group mobility model for ad hoc wireless networks,"* in *Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, 1999, pp. 53–60.

[9] M. Grossglauser and D. Tse, "*Mobility increases the capacity of ad-hoc wireless networks,"* in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies.Proceedings. IEEE*, 2001, vol. 3, pp. 1360–1369.

[10] Q. Dong and W. Dargie, "A survey on mobility and mobility-aware MAC protocols in wireless sensor networks," *Communications Surveys & Tutorials, IEEE,* vol. 15, no. 1, pp. 88–100, 2013.

[11] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira Jr, and C. Ratti, "Understanding individual mobility patterns from urban sensing data: A mobile phone trace example," *Transportation research part C: emerging technologies*, vol. 26, pp. 301–313, 2013.

[12] R. Becker, R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, and C. Volinsky, "*Human mobility characterization from cellular network data*," *Communications of the ACM*, vol. 56, no. 1, pp. 74–82, 2013.

[13] L. Pelusi, A. Passarella, and M. Conti, "*Opportunistic networking: data forwarding in disconnected mobile ad hoc networks*," *Communications Magazine, IEEE*, vol. 44, no. 11, pp. 134–141, 2006.

[14] R. J. D'souza and J. Jose, "*Routing approaches in delay tolerant networks: A survey," International Journal of Computer Applications, vol.* 1, no. 17, pp. 8–14, 2010.

[15] Q. Li, S. Zhu, and G. Cao, "*Routing in socially selfish delay tolerant networks*," in *INFOCOM, 2010 Proceedings IEEE*, 2010, pp. 1–9.

[16] Y. Li, G. Su, D. O. Wu, D. Jin, L. Su, and L. Zeng, "*The impact of node selfishness on multicasting in delay tolerant networks*," *Vehicular Technology*, *IEEE Transactions on*, vol. 60, no. 5, pp. 2224–2238, 2011.

[17] B. Jedari and F. Xia, "*A survey on routing and data dissemination in opportunistic mobile social networks*," IEEE Communications Surveys and Tutorials, to appear.

[18] N. Vastardis, K. Yang. "*Mobile Social Networks: Architectures, Social Properties and Key Research Challenges*", IEEE Communications Surveys and Tutorials. Volume: PP , Issue: 99, Pages:1-17, 2012.

[19] Y. Zhu, B. Xu, X. Shi, and Y. Wang, "*A survey of social-based routing in delay tolerant networks: positive and negative social effects*," *Communications Surveys & Tutorials*, *IEEE*, vol. 15, no. 1, pp. 387–401, 2013.

[20] P. Grindrod and D. J. Higham, *"A matrix iteration for dynamic network summaries," SIAM Review*, vol. 55, no. 1, pp. 118–128, 2013.

[21] 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.

[22] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "*Impact of human mobility on opportunistic forwarding algorithms*," *Mobile Computing, IEEE Transactions on*, vol. 6, no. 6, pp. 606–620, 2007.

[23] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, *Crawdad data set cambridge/haggle (v. 2006-01-31)(January 2006).*