Temporal Centrality Prediction in Opportunistic Mobile Social Networks

Huan Zhou[1], Shouzhi Xu[1(✉)], and Chungming Huang[2]

[1] College of Computer and Information Technology,
China Three Gorges University, Yichang, China
zhouhuan117@gmail.com, xsz@ctgu.edu.cn
[2] Department of Computer Science and Information Engineering,
National Cheng Kung University, Tainan, Taiwan
huangcm@locust.csie.ncku.edu.tw

Abstract. In this paper, we focus on predicting nodes' future importance under three important metrics, namely betweenness, and closeness centrality, using real mobility traces in Opportunistic Mobile Social Networks (OMSNs). Through real trace-driven simulations, we find that nodes' importance is highly predictable due to natural social behaviour of human. Then, based on the observations in the simulation, we design several reasonable prediction methods to predict nodes' future temporal centrality. Finally, extensive real trace-driven simulations are conducted to evaluate the performance of our proposed methods. The results show that the Recent Uniform Average method performs best when predicting the future Betweenness centrality, and the Periodical Average Method performs best when predicting the future Closeness centrality in the MIT Reality trace. Moreover, the Recent Uniform Average method performs best in the Infocom 06 trace.

1 Introduction

Recently, with the rapid proliferation of wireless portable devices (e.g., smartphones, ipad, PDAs) with bluetooth or wi-fi, Opportunistic Mobile Social Networks (OMSNs) begin to emerge [1–3]. Opportunistic Mobile Social Networks combine opportunistic networks and social networks together. Previous studies have shown that the performance of such networks depends highly on the user's social behavior as opportunistic networks and social networks share many common characteristics. Their common features motivate an increasing research interests in OMSNs, especially using the social network analysis technology to help the design of routing protocols [4].

Previous studies in OMSNs have proposed diverse metrics to measure the relative importance of a node in networks such as betweenness centrality, and closeness centrality [5]. However, when calculating such centrality metrics, the current studies focused on analyzing static networks that do not change over time or using aggregated contact information over a period of time. Actually, nodes in OMSNs are inherently dynamic, which is driven by natural social behaviour

of human. Therefore, it is not prudent to assume stationary human behavior in the design of practical applications. In particular, many researchers have also observed that nodes in OMSNs have a regular pattern [6,9]. For example, a student will go to school with his neighbours at morning every day, and have classes with his classmates in the classroom, which brings a regular pattern of physical contacts, etc., which in turn provides the periodicity seen in the underlying communication processes. Here, we make a simple assumption: since a node's schedule is regular, if it is an important node in the network at current time, then it is highly possible that its importance in the future will be correlated with the importance at current time.

In this paper, we therefore focus on predicting nodes' future importance under two important metrics, namely betweenness, and closeness centrality, using different real mobility traces in OMSNs. Actually, some studies have attempted to capture the dynamic behaviour of human. Tang et al. [7] proposed temporal centrality metrics based on temporal paths in order to measure the importance of a node in a dynamic network. Kim et al. [8] proposed several methods to predict nodes' importance in the future. Our work is similar to the work in [8], but we ignore the lagged time, and only predict the centrality value of a single time window in the network model, which makes the problem more clear and the prediction results more reasonable. Our contributions in this paper are three-folds:

1. Through real trace-driven simulations, we show that nodes' centrality values are highly correlated with their past recent centrality values, and also have periodical behavior at 24 h difference.
2. Based on the observations, we design several intuitive prediction methods, to predict nodes' future temporal centrality.
3. We evaluate the performance of the proposed prediction methods using different real mobility traces, and show that the best-performing prediction functions are more accurate on average than just using the last centrality value.

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

2   Preliminaries

In this section, we first introduce the network model related to this paper, and then define the notation and terminology for centrality metrics. Finally, we introduce the generalized network centrality prediction problem which will be used in the rest of the paper.

2.1   Network Model

We assume that the time during which a network is observed is finite, from t_{start} until t_{end}; without loss of generality, we set $t_{start} = 0$ and $t_{end} = T$. A dynamic network contact graph $G_{0,T} = (V, E_{0,T})$ on a time interval $[0, T]$ consists of a set of vertices V and a set of temporal edges $E_{0,T}$, where stochastic contact process between a node pair $u, v \in V$ on a time interval $[t_s, t_e]$ $(0 \leq t_s \leq t_e \leq T)$ is modeled as an temporal edge $e_{t_s,t_e}^{uv} \in E_{0,T}$.

   We divide the time interval $[0, T]$ into fixed discrete time windows $\{1, 2, ..., n\}$. We use $w = \frac{T}{n}$ to denote the size of each time window, expressed in some time unites (e.g., minutes or hours). In other words, a dynamic network can be represented as a series of static graphs at each time, $G_1, G_2, \ldots, G_n$. The notation G_t $(1 \leq t \leq n)$ represents the aggregate graph which consists of a set of vertices V and a set of edges E_t where an edge $e_{t_s,t_e}^{uv} \in E_t$ exists only if a temporal edge $e_{t_s,t_e}^{uv} \in E_{0,T}$ exists between vertices u and v on a time interval such that $t_e \leq wt$ and $t_s > w(t-1)$.

2.2   Network Centrality Measures

Centrality refers to a group of metrics that aim to quantify the "importance" or "influence" of a particular node (or group) within a network. There are several common methods to measure "centrality". In this paper, we only introduce two of them: betweenness, and closeness centrality. Formally, we use the standard definition of the betweenness, and closeness centrality, and the centrality value of a node i can be expressed as follows [5]:

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

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

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

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

$$Closeness(i) \; = \; \frac{1}{\mid V \mid -1} \sum_{j \neq i, j \in V}^{r} \Delta_{i,j}(i) \tag{2}$$

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

2.3 Centrality Prediction Problem

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

$$Error(G_t) \; = \; \frac{\sum_{i \in V} \mid C_t(i) - \hat{C}_t(i) \mid}{\mid V \mid} \tag{3}$$

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

3 Centrality Prediction

In this section, we first use two real mobility traces, *Infocom 06* [10] and *MIT Reality* [11] to test whether the centrality can be predicted. Then, based on the findings, we introduce several methods to predict the future temporal centrality.

3.1 Analysis of Correlation Between Past and Future Centrality

Note that human in reality always have regularity. Therefore, we hypothesize that the past centrality has high correlation with the future centrality. To test
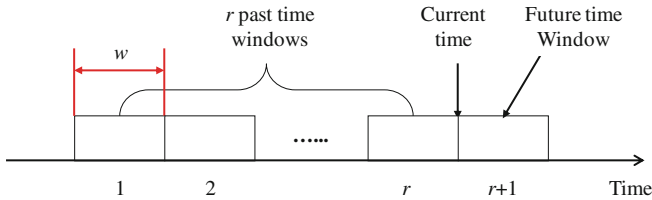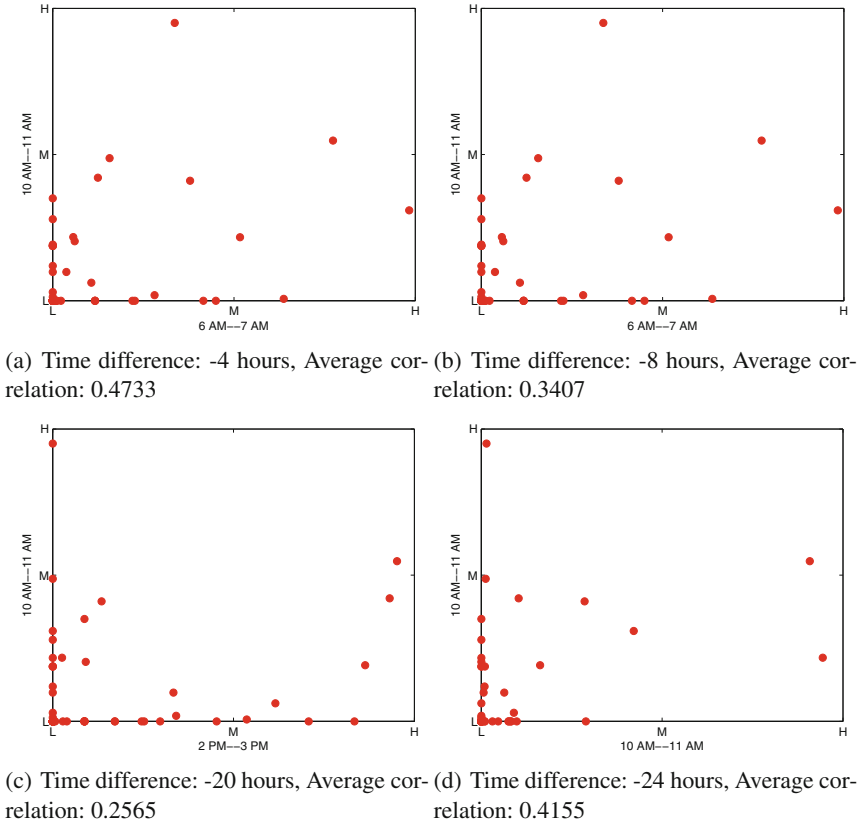

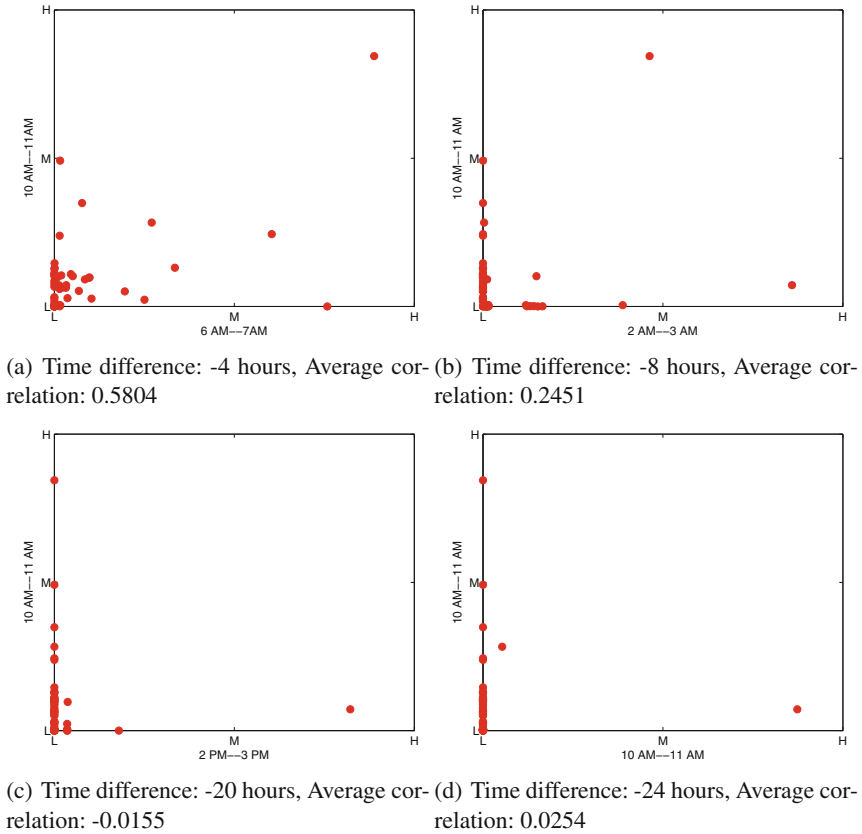**Fig. 1.** Illustration of the past and future time windows.

this hypothesis, we use two real mobility traces, *Infocom 06* and *MIT Reality* collected from real environments. Users in these two traces are all carrying Bluetooth-enabled portable devices, which record contacts by periodically detecting their peers nearby. The traces cover various types of corporate environments and have various experiment periods. For simplicity, we only analyze the correlation of Betweenness Centrality between the past and future time intervals.

(a) Time difference: -4 hours, Average correlation: 0.4733

(b) Time difference: -8 hours, Average correlation: 0.3407

(c) Time difference: -20 hours, Average correlation: 0.2565

(d) Time difference: -24 hours, Average correlation: 0.4155

Fig. 2. Scatter plots depicting correlation between a fixed window (y-axis) and an increasingly distant window from the past (x-axis) every 4 h in the MIT trace

Figure 2 plots each nodes' Betweenness centrality value compared to its value in a past window, in the MIT reality trace. It can be found that there is high correlation (0.4733) between a node's temporal centrality value with its value 4 h ago; second, increasing the time difference decreases the correlation (e.g., 0.3407 at -8 h difference); and third, at 24 h difference the correlation rises again (0.432), which indicates possible periodic behaviour.

Figure 3 shows each nodes' Betweenness centrality value compared to its value in a past window, in the infocom 06 trace. It can be found that similar to the

(a) Time difference: -4 hours, Average correlation: 0.5804

(b) Time difference: -8 hours, Average correlation: 0.2451

(c) Time difference: -20 hours, Average correlation: -0.0155

(d) Time difference: -24 hours, Average correlation: 0.0254

Fig. 3. Scatter plots depicting correlation between a fixed window (y-axis) and an increasingly distant window from the past (x-axis) every 4 h in the Infocom 06 trace

results in the MIT Realty trace, recent past temporal centrality values are highly correlated compared with more distant values(e.g., 0.5804 at -4 h difference). However, different from the results in MIT Realty trace, the pattern of repeated peaks with -24 h time difference seems rather weak in Infocom 06 trace. The main reason is that people attending the IEEE Infocom 2006 conference are more likely to seek out new colleagues to talk to at the breaks between sessions, rather than socialising with the same people, but students in the MIT campus are more likely to meet the same people when they are taking classes or walking in the campus.

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

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

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

3.2 Prediction Methods

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

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

Recent Uniform Average Method. In order to improve the accuracy of the prediction, we can use the node's m previous centrality values instead of one last previous centrality value. A reasonable idea is to use the node's uniform average centrality value between G_{r-m+1}, ..., G_{r-1}, G_r where $0 < m \le r$ as the node's future temporal centrality value. Formally, the future temporal centrality $C_{fu}(i)$ can be expressed as:

$$C_{fu}(i) \;=\; \frac{1}{m} \sum_{k=r-m+1}^{r} C_k(i) \tag{4}$$

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

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

$$f(k) \;=\; \begin{cases} 1, & kw \bmod p \equiv a, \\ 0, & kw \bmod p \neq a. \end{cases} \tag{5}$$

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

$$C_{fu}(i) \;=\; \frac{\sum_{k=r-m+1}^{r} f(k)C_k(i)}{\sum_{k=r-m+1}^{r} f(k)} \tag{6}$$

4   Performance Evaluation of Prediction Functions

In this section, we aim to evaluate the performance of the proposed methods above in the Infocom 06 and MIT Reality traces, and find the proper parameter values (e.g., m) of each prediction method at the same time. For each prediction method, as introduced above, we use $Error(G_t)$ to evaluate the prediction accuracy of the proposed methods.

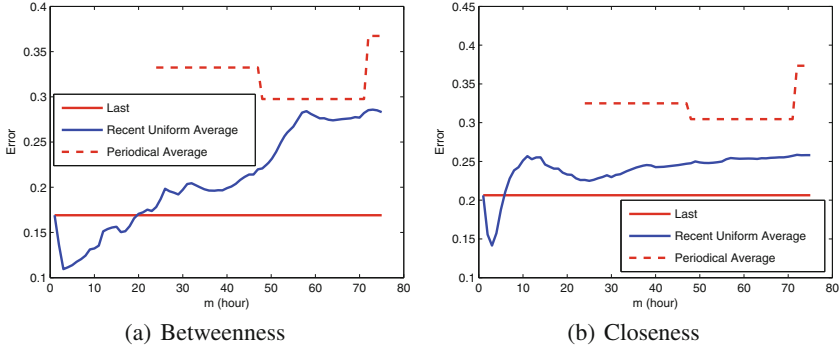(a) Betweenness   (b) Closeness
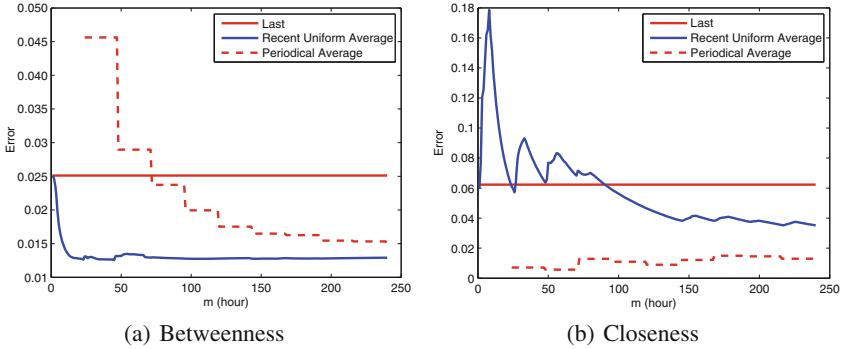
Fig. 4. The temporal centrality prediction results by varying m in the MIT Reality traces when w is 1 h

Figure 4 shows the temporal centrality prediction results of different prediction methods in the MIT Reality trace. It can be found that the Recent Uniform Average Method achieves the best results in Betweenness, and the Periodical Average Method achieves the best results in Closeness. Therefore, we recommend to use Recent Uniform Average method to predict the future Betweenness centrality, and use the Periodical Average Method to predict the future Closeness centrality in the MIT Reality trace. It is worth noticing that with the increasing of m, the Recent Uniform Average method achieves the minimum $Error(G_t)$ (0.01254) when m is around 45 h in Betweenness, and Periodical Average Method achieves the minimum $Error(G_t)$ (0.00569) when m is 49 h in Closeness. We would not recommend using Last method because its relative accuracy is not enough, although its computation cost is relatively cheap.

Figure 5 shows the temporal centrality prediction results of different prediction methods in the Infocom 06 trace. By contrast, it can be found that with the increasing of m, the Periodical Average method is not as good as that in the MIT Reality trace. Actually, we have already observed that there is no noticeable periodic patterns while the recent past centrality values are highly correlated in the Infocom 06 trace - Fig. 3 illustrates this. Therefore, we recommend to use Recent Uniform Average method to predict the future temporal centrality value in the Infocom 06 trace. Furthermore, the Recent Uniform Average method achieves the minimum $Error(G_t)$ (0.1095 in Betweenness, and 0.1415 in Closeness) when m is 3 h, and then the $Error(G_t)$ values will increase after this time interval.

Fig. 5. The temporal centrality prediction results by varying m in the Infocom 06 traces when w is $1\,\mathrm{h}$

This imply that the centrality value at a specific time in the Infocom 06 trace is highly related to the recent past centrality values in $3\,\mathrm{h}$.

In summary, we recommend to use the Recent Uniform Average method performs best to predict the future Betweenness centrality, and the Periodical Average Method to predict the future Closeness centrality in the MIT Reality trace, while we recommend to use the Recent Uniform Average method to predict the future temporal centrality in the Infocom 06 trace.

5   Conclusions

In this paper, we have predicted nodes' future importance under two important metrics, namely betweenness, and closeness centrality, using real mobility traces in OMSNs. Through real trace-driven simulations, we find that nodes' centrality value are highly correlated with their past recent centrality values, and have periodical behavior at $24\,\mathrm{h}$ difference. Then, based on the observations in the simulation, we design several reasonable prediction methods to predict nodes' future temporal centrality. Extensive real trace-driven simulation results show that the Recent Uniform Average method performs best when predicting the future Betweenness centrality, and the Periodical Average Method performs best when predicting the future Closeness centrality in the MIT Reality trace. Moreover, the Recent Uniform Average method performs best in the Infocom 06 trace.

References

1. Fan, J., Chen, J., Du, Y., Wang, P., Sun, Y.: DelQue: A socially-aware delegation query scheme in delay tolerant networks. IEEE Trans. Veh. Technol. **60**(5), 2181–2193 (2011)
2. Li, F., Wu, J.: MOPS: Providing content-based service in disruption-tolerant networks. In: IEEE ICDCS (2009)
3. Zhou, H., Zhao, H., Chen, J.: Energy saving and network connectivity tradeoff in opportunistic mobile networks. In: IEEE Globecom (2012)
4. Zhou, H., Chen, J., Zhao, H., Gao, W., Cheng, P.: On exploiting contact patterns for data forwarding in duty-cycle opportunistic mobile networks. IEEE Trans. Veh. Technol. **62**(9), 4629–4642 (2013)
5. Scott, J.: Social network analysis. Sociol. **22**(1), 109–127 (1988)
6. Zhou, H., Chen, J., Fan, J., Du, Y., Das, S.K.: ConSub: incentive-based content subscribing in selfish opportunistic mobile networks. IEEE J. Sel. Areas Commun. **31**(9), 669–679 (2013)
7. Tang, J., Kim, H., Anderson, R.: Temporal node centrality in complex networks. Phys. Rev. E **85**(2), 2181–2193 (2012)
8. Kim, H., Tang, J., Anderson, R., Mascolo, C.: Centrality prediction in dynamic human contact networks. Comput. Netw. **56**(3), 983–996 (2012)
9. Gao, W., Li, Q., Zhao, B., Cao, G.: Multicasting in delay tolerant networks: a social network perspective. In: Mobihoc, pp. 299–308. ACM (2009)
10. Scott, J., Gass, R., Crowcroft, J., Hui, P., Diot, C., Chaintreau, A.: CRAWDAD data set cambridge/haggle (v. 2009–05–29) (2009). http://crawdad.cs.dartmouth.edu/cambridge/haggle
11. Eagle, N., Pentland, A.S., Lazer, D.: Inferring friendship network structure by using mobile phone data. Proc. Nat. Acad. Sci. **106**(36), 15274–15278 (2009)