# Exploiting Time-varying Graphs for Data Forwarding in Mobile Social Delay-Tolerant Networks

Dongliang Xie[2,1], Xin Wang[1], Lanchao Liu[2], Linhui Ma[2]
[1] Department of Electrical and Computer Engineering, State University of New York at Stony Brook, USA
[2] State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications, China
dongliang.xie@gmail.com, xwang@ece.sunysb.edu, LanchaoLiu92@gmail.com, malinhui@bupt.edu.cn

***Abstract*—With the rapid shift from end-to-end communications to content-based data sharing, there are increasing interests in exploiting mobile social Delay-Tolerant Networks (social DTNs) to deliver data, where the forwarding decision is usually made by comparing the social metrics of encountered nodes. Existing studies mostly derive long-term statistical social metrics without considering the temporal impact from node mobility.**

**We exploit the time-varying contact graphs to analyze the dynamics of social DTNs based on two groups of datasets. Based on the analysis, we derive the time-varying characteristics of node contacts, *durative* and *periodicity*, and apply them to more accurately predict the corresponding time-varying social metrics (TSMs). We further propose a two-stage opportunistic forwarding strategy to select relays based on TSMs. Our simulation results verify the importance of the two properties we observe and the effectiveness of our algorithm in tracking time-varying social metrics. We also show the potential of our algorithm in finding general time varying metrics to improve the data dissemination performance of other opportunistic forwarding schemes.**

## I. INTRODUCTION

With the rapid growth of content-based data sharing applications and the popularity of mobile devices, there are increasing interests in exploiting mobile social Delay-Tolerant Networks (social DTNs) [9] to deliver data without relying on a well-established network infrastructure [31]. A social DTN is characterized with intermittent connections and opportunistic node encountering, where a data carrier can select an appropriate relay to continue forwarding the data towards the destination [22][28].

The limited connection in DTNs results in the lack of stable end-to-end paths, large transmission delay and unstable network topology. The store-carry-forward pattern to transfer data can be used in social DTNs,and it will get good performance if we choose suitable next-hop relay. The key challenge is how to select the best next-hop relay. In general,we can select a proper relay by some appropriate metrics. Metrics such as the probability of data delivery [3] and the number of historical contacts [4][30] have been proposed in the literature to achieve more reliable data delivery. These studies, however, did not consider social relationship of nodes. To more accurately choose the next-hop relay, social characteristics such as *centrality* and *similarity* [18] are recently applied in the forwarding decision process [1], [6], while a social contact graph is used to derive social metrics in [13], [23]. However,there are some deviations between the research and the reality because of the characteristic of DTNs.

The social metrics derived based on long-term statistical social features did not take into account the impact of node mobility on the constant changes of the social relationship, and the inaccurate social metrics may result in wrong relay selection thus poor performance in the data forwarding. The initial attempts [21], [7] to address this issue, although important, took over-simplified social metrics, and also did not provide algorithms to effectively track the evolving metrics.

In this paper, we propose a novel scheme to efficiently support real-time opportunistic forwarding in social DTNs. We first derive important time-varying characteristics of node behaviors by analyzing two sets of social data, based on which we provide an efficient algorithm to predict time-varying social metrics (TSMs). We also propose a novel opportunistic forwarding mechanism that concurrently consider three types of time-varying social metrics: *betweenness centrality* [14], *similarity* and *tie strength* [19].

Our contributions can be summarized as follows:

- We first exploit the *time-varying contact graphs* to analyze the dynamics of social DTNs, and demonstrate its superiority over traditional statistical data analysis in capturing the time-varying features of social DTNs.
- We derive two important properties of time-varying social contacts from data analysis, *durative* (either *continuous* or *discrete*) and *periodicity* , and present a *window-based* method to capture the impact of the time-varying characteristics on future contact behaviors. Based on the properties observed and the analysis of *ranking correlation* of different windows, we propose an algorithm to more accurately predict the social metrics over time.
- We propose a two-stage opportunistic packet forwarding mechanism to fully exploit all the time-varying social metrics proposed for higher forwarding performance.
- We evaluate the performance through extensive simulations. Our results verify the effectiveness of our proposed time-varying social metrics and prediction/forwarding

algorithms in achieving higher data forwarding performance.

Our proposed approach for calculating the time-varying social metrics is general and not constrained to the metrics we have derived, and can be applied to work with other existing forwarding strategies. In addition, the properties of node behaviors observed and verified can serve a guidance for more efficient designs of algorithms for social networks.

The rest of this paper is organized as follows. Section II reviews the related work. Section III gives an overview on the *contact-compare-forward* method for opportunistic forwarding and the social metrics we consider. We analyze the time-varying characteristics of social contacts and social metrics in real social networks in Section IV. In Section V, we provide a scheme to effectively predict time-varying social metrics and a strategy to enable more efficient opportunistic forwarding based on TSMs. In Section VI, we evaluate the performance of our forwarding strategy and TSMs prediction method through extensive simulations based on two groups of trace data. Finally, we conclude our work in Section VII.

## II. Related Work

The development of DTNs routing protocols has undergone several phases, including random multi-copy forwarding, forwarding based on metric comparison, and opportunistic forwarding based on social metric comparison. In the early stage DTNs routing, Epidemic [3] employs flooding to distribute data in the network where nodes freely copy their data to ones they encounter. Spray and Wait [4] improves Epidemic by restricting the number of copies a node makes when forwarding the packet to reduce the consumption of network resources. The transmissions of multiple copies help increase the packet delivery ratio at the cost of high network resource consumption. However, the blind and random packet forwarding makes the transmission efficiency very low.

Subsequently, a class of forwarding mechanisms based on metric comparison are applied. ProPHET [5] and MaxProp [6] are two of such mechanisms which assume that the movement of nodes is not purely random and estimates the delivery probability to the destination based on the historical contact information. The delivery probability is characterized as the comparison metric for making forwarding decision. The use of comparison metric avoids the transmission blindness in the traditional multi-copy based on forwarding mechanism. In the practical social networks, regularity and randomness coexist due to node mobility. The assumption that the nodes' movement is absolutely not random without considering their social attributes may cause an inaccurate determination of the forwarding metrics.

Social metrics are generally classified into two types, global and local. In Social Network Analysis (SNA), the global node attribute *centrality* is defined based on the difference in physical significance of nodes. Authors in [16] further classify the *centrality* into three types: *degree centrality*, *closeness centrality*, and *betweenness centrality*. Defined based on the number of connecting neighbors of a node, the *degree centrality* however only reflects the node's local structure [17]. Instead, the *Closeness centrality* of a node represents how fast it reaches all the other nodes, while the *betweenness centrality* reflects the extent for a node to locate at the shortest paths between all other nodes.

There are two major types of local metrics. The *similarity* is determined based on the number of common neighbors between a node pair [18]. The *tie strength* [19] reflects the mutual relationship between two nodes, and is characterized in terms of *the amount of time, the emotional intensity, the intimacy (mutual confiding)*, and *the reciprocal services.*

SimBet [1] and BubbleRap [6] are the typical opportunistic forwarding mechanisms that employ social metrics to determine the next hop. In SimBet, the forwarding decision is made based on a utility, which combines the *centrality* and *similarity* calculated using the the static contact graph derived from the historical statistics of contacts. In a practical opportunistic network, the social status and relationship of nodes will change over time due to the mobility, joining and departure of nodes. Such changes make the social metrics calculated based on long-term historical data inaccurate. The subsequent version SimbetTS [5] adds the *tie strength* for utility calculation, while the inaccuracy due to the dynamics in social relationship still exists. BubbleRap divides the *centrality* into global and local to apply outside and inside the community respectively, but still does not adapt to the network dynamics.

SimBetAge [21] improves upon SimBet by adopting an aged graph to calculate the social metrics dynamically. The aged graph only uses time as the weight, but did not consider the regularity characteristics of contacts inherent in many social applications, which impacts the accuracy of the metric thus the network performance. Transient [7] calculates the *centrality* based on the transient contact pattern of nodes, and applies it in a forwarding strategy similar to that of BubbleRap. The centrality only reflects the direct and indirect relationship between neighboring nodes rather than the relationship between nodes and the destination, which may lead to an inefficient relay selection or no chance of forwarding data to the destination. In contrast, our proposed TSMs exploit metrics that better represent the relationship between candidate relays and the destination. Both SimBetAge and Transient simply compare the metrics derived without prediction of metrics when the time evolves. As only the *centrality* metric is considered, the performance of Transient is compromised.

Different from the wide diversity of the literature work which directly uses the social metrics calculated, the observations of the *durative* (i.e., continuous and discrete) and *periodicity* node behaviors allow us to more accurately predict the social metrics over time. We also propose a two-stage time-dependent opportunistic forwarding mechanism in social DTNs by exploiting both the global and local time-varying social metrics predicted. Our method used in the analysis and exploration of *time-varying contact graphs* is fundamental and general, which can be extended to derive other time varying metrics of social networks.

In this paper, we exploit the *time-varying contact graphs* to analyze the dynamics of social DTNs, and derive the rules and the special time-varying characteristics of the social networks. Different from the wide diversity of the literature work which directly uses the social metrics calculated, the observations of the *durative* and *periodicity* features allow us to more accurately predict the social metrics over time.

## III. Network Model and Metrics

We first introduce the basic forwarding process, and then the social metrics to use for data forwarding in social DTNs.

### A. Basic Forwarding Mechanisms in Social DTNs

Opportunistic forwarding mechanisms normally follow a *contact-compare-forward* pattern. When a node $i$ with data has a contact with the node $j$, if $j$ is the destination, $i$ forwards data directly to $j$ and completes the transmission; otherwise, $i$ needs to compare its routing metric with $j$ to determine if selecting $j$ as the relay to forward its data to the destination.

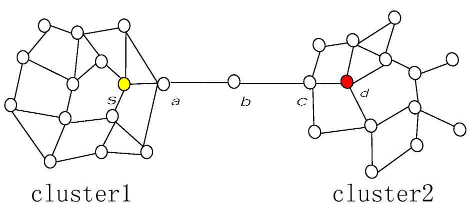



Fig. 1: Opportunistic forwarding based on social metrics

Most of existing schemes select relays based on both the global and local metrics. In Fig. 1, a source $s$ from the cluster 1 wants to send data to a destination $d$ in cluster 2. As $a$ is able to communicate with nodes outside the cluster and thus more important globally than other neighbors, $s$ forwards data to $a$ in the first step. Nodes $a$ and $b$ have approximately equal global importance due to their positions in the network, so their local metrics are compared. Obviously $b$ is closer to the destination than $a$ , so data should be delivered to $b$ to continue the forwarding towards $d$.

### B. Social Metrics

As delay tolerant data transmissions are often needed in applications involving social relationship, it helps to consider the characteristics of social networks. The metrics we take also include two categories of metrics: global and local. In social networks, some nodes have higher chance of meeting others because of their special social status, so these nodes are considered as the ones that have higher global significance or centrality. However, if we only take into account the global metric in the transmission process, data may be only passed among nodes with higher global metrics which may be far from the destination. Therefore, we need to also apply local metrics to measure nodes' social relationships. The local attributes can be reflected in the real world, for example, one node may maintain different social ties with others according to nodes' social relations, daily habits and work styles. Particularly, a node with a closer relation to the destination usually accomplishes the data delivery faster than the ones with higher global popularity.

*1) Global Metrics:* We select the *betweenness centrality* as the global metric in our forwarding mechanism, as this metric can reflect a node's capability of controlling the communications between other nodes. It is more appropriate to guide the data transmissions in social DTNs than *degree centrality* and *closeness centrality* discussed in Section II.

In [25][26], the *betweenness centrality* is defined as

$$Bet_v = \sum_{i}^{n} \sum_{<\ j}^{n} b_{ij}(v), \tag{1}$$

where $b_{ij}(v)$ represents the fraction of the number of shortest paths between $i$ and $j$ that go through node $v$ among all the shortest paths of $(i, j)$. In this paper, we also call *betweenness centrality* as *betweenness*.

*2) Local Metrics:* For local metrics, we adopt *similarity* and *tie strength*. *Similarity* is defined as

$$S_v(d) = |N(v) \cap N(d)|, \tag{2}$$

which represents the number of common neighbors of a node pair $(v, d)$ and reflects the indirect relationship between nodes.

*Tie strength* [5] [19] measures the strength of the relationship between a node pair. In this paper we use the proportion of the contact duration $(v, d)$ in a time period $T$ to measure the strength of the relationship between two nodes $v$ and $d$:

$$TS_v(d) = \frac{d(v, d)}{T}. \tag{3}$$

As *similarity* and *tie strength* are local metrics, they can be calculated using the local graph conveniently. However, it is difficult to obtain the *betweenness* which requires the global contact information and the global contact graph to derive. In the ego network [11], local contact graphs and social metrics are built based on local records and the interactions among neighboring nodes. In [12], the neighbor matrix is applied to calculate the *ego betweenness*, and the ranking of *ego betweeness* among nodes is shown to be the same as the ranking of the global *betweenness*. Thus, all social metrics can be gained based on the local contact graph of nodes.

## IV. Time-varying Characteristics Analysis

In this section, we introduce basic features of the social contacts and verify the features through two sets of social data collected from the practical world. Specifically, we will show the time-varying features of the social contacts and derive the characteristics of social metrics based on the features.

### A. Social Contact Graph

In this subsection, we compare the social metric calculation based on statistical contact graph and time-varying graph.

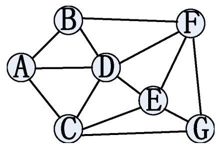


Fig. 2: Static statistical contact graph $G$

TABLE I: Betweeness values in static statistical contact graph

| NodeID | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| Betweenness | 0.5 | 0.5 | 1.833 | 2.333 | 0.667 | 1.833 | 0.333 |

*1) Static Statistical Contact Graph:* Based on the static statistical data analysis, the social metrics mentioned in Section III can be calculated from a contact graph. For example, with the contact graph shown in Fig. 2, the *betweenness* value of each node is calculated and shown in Table I.

Obviously, the *betweenness* value of node $D$ is the highest, the values of $C$, $F$ are slightly lower, while $G$ has the lowest value. However, the *betweenness* of nodes may change over time as a result of the edge evolution. Assume that nodes in the contact graph have different connections at different time of a day, we can decompose a contact graph into three sub-graphs:

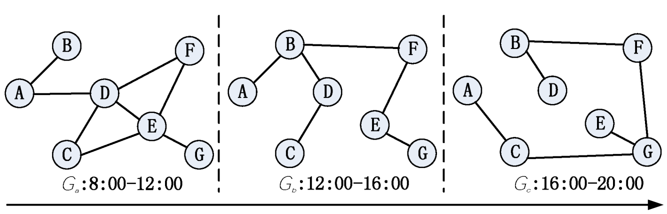



Fig. 3: Time-varying contact graphs of $G$

We calculate the *betweenness* values from $G_a$ to $G_c$ and show the results in Table II.

TABLE II: Betweeness values in time-varying contact graphs

| NodeID | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| Betweenness($G_a$) | 5 | 0 | 0 | 8.5 | 5.5 | 0 | 0 |
| Betweenness($G_b$) | 0 | 11 | 0 | 5 | 5 | 8 | 0 |
| Betweenness($G_c$) | 0 | 4 | 5 | 0 | 0 | 8 | 11 |

There exists a big difference between the values shown in Table I and Table II. In the graph $G_a$, the *betweenness* of node $D$ is the largest but is *zero* in the graph $G_c$. Node $C$ has the third largest *betweenness* in $G_c$ but zero *betweenness* in graphs $G_a$ and $G_b$. Although the *betweenness* of node $G$ is the minimum in the contact graphs $G_a$ and $G_b$, it becomes the maximum in the graph $G_c$. These data indicate that the candidate relay nodes may change over time.

*2) Time-Varying Graph:* In order to analyze the time-varying characteristics of social DTNs, we first describe the network in the form of time-varying graph [8]. In this paper, a time-varying graph is defined as $G \equiv (V, E, L, \omega)$, where $V$ is the vertex set formed with mobile nodes in the network, $E$ is the set of edges and each edge represents a contact between nodes, $L$ is the lifetime of messages in a social network beyond which the undelivered messages will be discarded, and $\omega$ is the time-window duration to determine the size of each sub-graph.

For ease of description, we define a static aggregate graph $G = (V, E)$, where $E$ includes all contacts in the duration $L$. With each snapshot being a contact graph of a time-window, a graph $G$ can be divided into multiple consecutive sub-graphs according to the time window sizes, namely a collection of time-varying graph snapshots: $G_1, G_2, ..., G_t, ...$:

$$G_{0,t}(V, E_{0,t}) = \{G_1(V, E_{0,\omega}), G_2(V, E_{\omega,2\omega}), \\ ..., G_t(V, E_{t-\omega,t})\}, \quad (4)$$

where $V$ is a set of nodes, $E_{t-\omega,t}$ represents the temporal contacts of nodes within the time window $(t-\omega, t)$, and $\omega$ is the time duration of each snapshot. The symbol $\omega$ represents time-window duration, and the time $t = n * \omega$.

Then we can map metrics defined in Equations (1) to (3) to those in a time window as

$$B_{v,t} = \sum_{i}^{n} \sum_{<}^{n} \sum_{j} b_{ij}^{(t-\omega,t)}(v), \quad (5)$$

$$S_{v,t}(d) = |N^{(t-\omega,t)}(v) \cap N^{(t-\omega,t)}(d)|, \quad (6)$$

$$TS_{v,t}(d) = \frac{d^{(t-\omega,t)}(v,d)}{\omega}, \quad (7)$$

$B_{v,t}$, $S_{v,t}(d)$ and $TS_{v,t}(d)$ are also called as *window metrics* whose values change in different time windows.

## *B. Time-Varying Characteristics of Social Contacts*

Based on the time-varying graph, we first analyze the dynamic characteristics of social contact, we then extract two new properties, *durative* and *periodicity*, and verify these properties through detailed analysis of real datasets.

*1) Properties of Time-varying Social Contacts:* Before we discuss the new properties of the time varying social contact, we give a definition below.

**Definition:** ***Recurrence of edges.*** In a time-varying graph $G \equiv (V, E, L, \omega)$, if for an edge $e \in E$ and $t \in L$, $\exists t' > t$: $e \in E_{t'-\omega,t'}$, the edge $e$ has the characteristic of recurrence.

This characteristic indicates that any edge $e$ in the network will reappear in a future time window $(t'-\omega, t')$. However, in the practical social networks, an edge $e$ reflects the contact between nodes. Not all edges are recurrent. For example, a node $i$ will leave the network forever after it contacts node $j$, then $e_{ij}$ will not recur. In addition, because information needs to be timely, assuming the edge will recur after an infinite time duration $\Delta t = t' - t$ also has no meaning. So, we will incorporate the social networks' characteristics into the definition, and we have the following two properties for social contacts.

**Property 1:** ***Durative of social contact.*** Assume $E_{t-\omega,t}$is a set of edges within a time-varying window $(t-\omega, t)$. If $\exists e_{i,j} \in E_{t-w,t} \bigcap E_{t,t+w}$, i.e., $e_{i,j}$ can recur in continuous snapshots before and after the time $t$, we call the contact of the node pair $(i, j)$ to be durative.

In social networks, contacts are not completely random due to the intrinsic social relationship among nodes. A certain pair of nodes may be classmates, work colleagues or conference participants. Based on ***Property 1***, durative can be categorized into two types, *continuous durative* and *discrete durative*. As an event is generally not instantaneous but can continue for a duration of time, the contact $e_{i,j}$ between a node pair $(i,j)$ will last and show the property of ***continuous durative***. It is more common that some events can occur again after a brief time interval due to the dynamic of DTNs. Take the conference scenario as an example, if a node leaves the network temporarily for an extremely short period of time (e.g. about 5 minutes), it should not impact the durative of a contact which lasts for a long period of time (e.g., one or more hours) [27]. Thus, if the contact $e_{i,j}$ recurs repeatedly within a period of time, it has the property of ***discrete durative***.

**Property 2:** ***Periodicity of social contacts.*** Assume $E_{t-\omega,t}$ is a set of edges within a time-varying window $(t-\omega,t)$, $T$ represents a time period which is usually set as $k\omega$, where $k \in N^+$. If $\exists e_{i,j} \in E_{t-\omega-T,t-T} \cap E_{t-\omega,t}$, then $e_{i,j}$ can recur in periodical windows and the contact of node pair $(i,j)$ has the periodicity feature. Obviously, when $k=1$, periodicity is equivalent to durative.

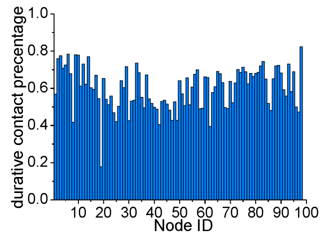
For any node $i$, its daily activities could follow some regularity. For instance, people could take a fixed set of buses to go to work, or have a meal in a fixed restaurant at some specific time. Therefore, a node $i$ might meet a node $k$ who has the same daily schedule. In this case, we consider the contact of $(i,k)$ reappears following a daily period. To more accurately predict the contact relationship between nodes, we should take into account both types of durative.

*2) Verifying the Properties of Social Contacts:* To verify the *durative* and *periodicity* of the social contacts, we choose two datasets which record nodes' behaviors in two practical social networks, Infocom06 collected from the conference environment and Cambridge collected from the university environment [10]. Both datasets capture contacts between nodes using Bluetooth devices, and the parameters are summarized in Table III.

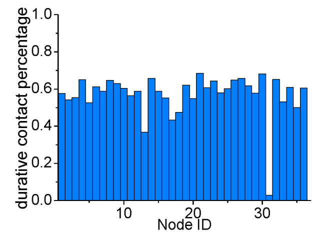
TABLE III: Specific information of two real datasets

| Dataset | Infocom06 | Cambridge |
|---|---|---|
| Device | IMote | IMote |
| Network type | Bluetooth | Bluetooth |
| Duration(days) | 3 | 11 |
| Granularity(Seconds) | 120 | 600 |
| Number of devices | 98 | 36 |
| Number of internal contacts | 141207 | 10641 |
| Number of node pairs | 8157 | 1033 |

The major difference of the two datasets is the count of the average number of contacts for each node pair per day. Constrained to the conference environment, nodes in Infocom06 have more frequent contacts than those of Cambridge. To derive some general rules followed by both, there is a need to properly choose the time window size $\omega$. Too large a window size cannot capture the time-varying features of contacts, while too small a window size may miss the contacts of discrete durative and lead to inaccurate calculation of social metrics. Generally, $\omega$ can be adjusted flexibly according to the number of nodes and the contact density. For both datasets, the duration of one hour appears to be common for durative and we verify this window size is reasonable in the Section VI-A through the simulation.

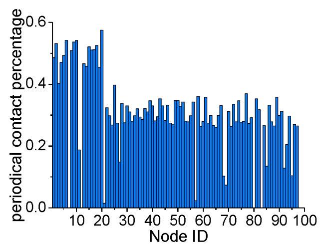


(a) Infocom06

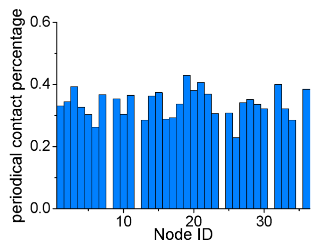


(b) Cambridge

Fig. 4: Durative contacts percentage of each node



(a) Infocom06



(b) Cambridge

Fig. 5: Periodicity contacts percentage of each node

Fig. 4 shows the time fraction of the durative contacts (including continuous and discrete ones) in all contacts with the time window set to $\omega = 1h$. With a restrictive activity space, for most nodes the durative percentages of Infocom06 are between 50% and 80%, The durative percentages in Cambridge are slightly lower, and are around 50% and 60%.

Fig. 5 shows the ratio of periodic contacts among all contacts for each node, with the time period set to $T = 24h$. The percentage of periodic contacts in each dataset is about 30%, and only some individual nodes have too high or too low periodicity. The reason that some nodes have too low contact periodicity is because they contact with others very little, or because the Bluetooth devices used to measure the social relationship are broken or forgotten to be carried by the people who participate in the measurement. The ones with too high contact periodicity are resulted from their relatively simple behaviors. For example, the number of nodes they contact is small and the nodes they contact at the same time of the day are quite fixed.

The above analyses of real social network datasets verify the durative and periodicity of nodes' contacts. We can apply them to explore time-varying characteristics of social metrics.

### C. Analysis of Time-Varying Characteristics of Social Metrics

Based on the exploitation of dynamic characteristics of social DTNs, we infer the feature of social metrics and present a *window-based* methodology to capture the impact of the time-varying characteristics on future contact behaviors.

*1) Properties of Social Metrics:* To facilitate opportunistic forwarding, we only care about the relative values of metrics and their comparison, rather than the absolute values. Social metrics are directly impacted by the interactions among different nodes. As the contacts of the node pair $(i, j)$ usually have the properties of *durative* and *periodicity*, the metrics of node $i$ and $j$ have similar time-varying characteristics.

Without loss of generality, three social metrics of node $v$ within time windows $(t-\omega, t)$ and $(t, t+\omega)$ are uniformly represented as $m_{v,t}$ and $m_{v,t+1}$. With a time period $T$ set as $T = k\omega$, where $k \in N^+$, $p_{v,t} = m_{v,t-T+1}$ represents the metrics of node $v$ within the time window $(t-T, t-T+\omega)$. We can derive the time-varying characteristics of the social metrics' based on ***Property 1*** and ***Property 2*** in Section IV-B:

**Lemma 1:** ***Durative of social metrics.*** For a node pair $(i, j)$, $\exists t \in T$, if $m_{i,t} > m_{j,t}$ and $m_{i,t+1} > m_{j,t+1}$, then we consider the metrics' ranking of $(i, j)$ remains the same before and after the time instant $t$.

**Lemma 2:** ***Periodicity of social metric.*** For a node pair $(i, j)$, $\exists t \in T$, if $p_{i,t} > p_{j,t}$ and $m_{i,t+1} > m_{j,t+1}$, then we consider the metrics' ranking of $(i, j)$ remains the same in the time windows with the time period $T$.

*2) Correlation Analysis of Window Metrics:* We use the real datasets to validate the time-varying characteristics of social metrics shown in ***Lemma 1*** and ***Lemma 2***.

Assume the current time is $t$, and we would like to know the node's capability of forwarding data within a future time window $W_{t+1}$, called a *prediction window*. In Fig. 6, to test the durative and periodicity of metrics, we calculate the correlation between $m_t$ of the window $W_t$ and $m_{t+1}$ of $W_{t+1}$ as well as the correlation between $m_{t-T+1}$ of a periodic window $W_p$ in the duration $(t-T,\ t-T+\omega)$ and $m_{t+1}$. We also choose the metrics within the second recent window $W_{t-1}$ and static metrics as the references. The static metrics $m_{(0,t)}$ are derived based on data accumulated within a long time duration $W_c$ from time 0 to $t$.

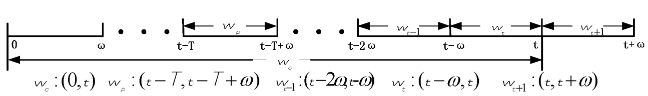
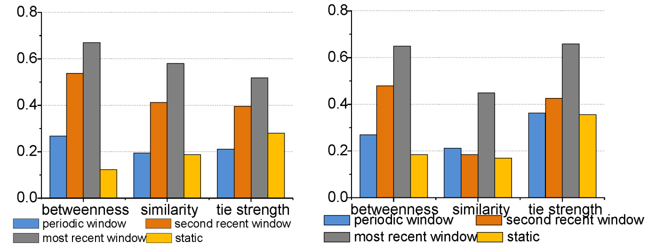



Fig. 6: Time-varying window partition

As only the relative metric ranking is important for the opportunistic forwarding process, we calculate the ranking correlation coefficient $\rho$ [15] [24] which reflects the correlation of metric ranking of all nodes:

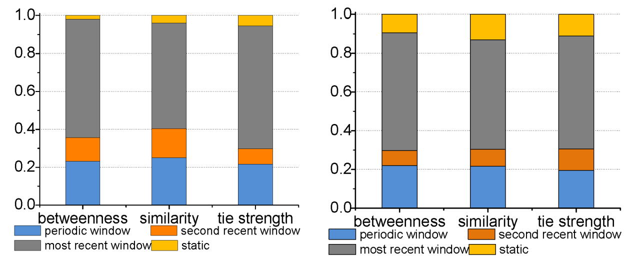
$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)},$$

where $n$ is the number of nodes in a dataset, and $d_i$ represents the difference of metric ranking between the prediction window $W_{t+1}$ and each of the following windows: $W_t$, $W_{t-1}$, $W_p$, $W_c$.




(a) Infocom06 (b) Cambridge

Fig. 7: Average ranking correlation of window metrics




(a) Infocom06 (b) Cambridge

Fig. 8: Highest ranking correlation percentage

In Fig. 7, with $\omega = 1h$ and $T = 24h$, we show the ranking correlation coefficient between the metric $m_{t+1}$ and each of the following metrics, $m_t$, $m_{t-1}$, $m_{t-T+1}$ and $m_{(0,t)}$ at different time instants $t$ with the same two datasets we used earlier. The local metrics, *similarity* and *tie strength* are calculated through a randomly chosen node with all other nodes.

Both datasets show the same trend. Among them, the average ranking correlation coefficient between the prediction window $W_{t+1}$ and the most recent window $W_t$ is the highest, and between $W_{t+1}$ and the second recent window $W_{t-1}$ is the second largest, which prove the durative of social metrics. The ranking correlation coefficient with the static window $W_c$ is the worst, which indicates that the static metrics can not well track the time-varying node relationship and using the static metric will reduce the forwarding efficiency and reliability. The ranking correlation coefficient of social metrics with $W_p$ is between those with $W_{t-1}$ and with $W_c$. The correlation of the static *tie strength* is higher than that with $W_p$ in Fig. 7(a), which may be due to the random selection of nodes in this metric calculation.

We show the percentage of time when different types of window has the highest ranking correlation coefficient in Fig. 8. Most of time, $W_t$ has the highest ranking correlation, and its percentages for all metrics in each dataset are over 50%∼60%. The second place is $W_p$ with the percentage around 20%∼30%. Which is a complementary to the result in the case of $W_t$ at some time points. This verifies the existence of periodicity in social metrics. Our analysis indicates that the recent window metric $m_t$ and the periodic window metric

$m_{t-T+1}$ contribute most to the prediction of TSMs in the future window, i.e., $m_{t+1} = f_1(m_t) + f_2(p_t)$. We thus exploit an improved Kalman Filer to predict future metrics.

## V. Forwarding Mechanism

In this section, we first present an analytical model to predict the temporal social metrics (TSMs) taking into account the features of social metrics, we then propose an opportunistic forwarding mechanism based on TSMs.

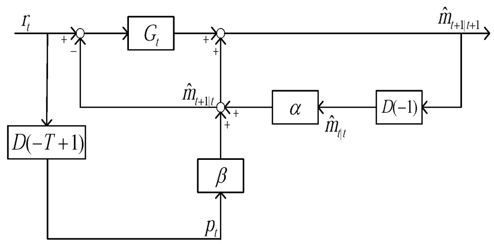


Fig. 9: Improved Kalman Filter for TSMs prediction process

### A. TSM Prediction Model

According to Section IV, the key problem of time-varying forwarding is to predict nodes' social metrics based on the durative and periodicity.

With the advantage of prediction accuracy and light-weight calculation, Kalman Filter (KF) is a good candidate tool for the prediction of time-varying social metrics in resource constrained mobile devices. For better prediction performance, we improve KF by integrating the metric values from the previous window and the periodic window as the input.

At the current time $t$, in order to predict the future TSMs value $m_{t+1}$ in the time window $(t, t+\omega)$, the previous window metric $m_t$ and the periodic window metric $m_{t-T+1}$ can be input as the measured values into KF. In the basic KF, the system states evolve only based on the recent window metric:

$$\begin{cases} m_{t+1} = A_t m_t + V_t \\ r_t = C_t m_t + W_t \end{cases} \tag{8}$$

where $r_t$ represents the social metric value observed in the window $(t-\omega, t)$. If we predict each metric independently, the vector of the system state metric $A_t$ is a scalar, and the state transition matrix $C_t = 1$. $V_t$ and $W_t$ are both the white Gaussian noise, where $Var(V_t) = Q_t$, $Var(W_t) = R_t$, and $E(V_t W_t^T) = 0$.

As shown in ***Lemma 2***, the social metrics have periodicity, so we add the periodic window metric value $p_t = r_{t-T+1}$ as a deterministic input of the system, then the metric prediction equation changes into

$$m_{t+1} = \alpha m_t + \beta p_t + V_t \tag{9}$$

where $\alpha + \beta = 1$.

With both the recent window metric $r_t$ and the periodic window metric $p_t$ as the input of the filter, the improved Kalman Filter is shown in Fig.10. The implementation of Kalman Filter has two stages: prediction and correction. To predict the metric value $m_{t+1|t+1}$ in $(t, t+\omega)$, we first give a prior prediction value $\hat{m}_{t+1|t}$ using both the prediction value $\hat{m}_{t|t}$ for the current window (performed during the previous window time) and the periodic window metric value $p_t$ based on Equation (9). We then modify the future metric $\hat{m}_{t+1|t+1}$ by comparing the observed recent window metric $r_t$ with the prediction value $\hat{m}_{t|t}$. The recursive formula is

$$\begin{cases} \hat{m}_{t+1|t} = \alpha \hat{m}_{t|t} + \beta p_t \\ \hat{m}_{t+1|t+1} = \hat{m}_{t+1|t} + G_t(r_t - \hat{m}_{t|t}) \end{cases} \tag{10}$$

where $G_t$ is the Kalman Gain for predicting $\hat{m}_{t+1|t+1}$ at the time $t$, which is given by Equations (11) and (12):

$$G_t = P_{t+1,t}(P_{t+1,t} + R_{t+1})^{-1} \tag{11}$$

$$P_{t+1,t} = \alpha P_{t,t} \alpha^T + Q_t \tag{12}$$

$P_{t+1,t}$ needs to be updated as follows:

$$P_{t+1,t+1} = (1 - G_t) P_{t+1,t} \tag{13}$$

At $t = 0$, we have the initial value $P_{0,0} = Var(m_0)$ and $m_{0|0} = E(m_0)$. For a certain time period right after the network initialization, the characteristics of social metrics are not stable. We consider a preparation stage at the beginning time of the network to obtain initial metric values, and set a time point after this stage to 0 as the initial time of the filter, where $m_0$ represents the metric value in $(-\omega, 0)$.

For the unified handling, we first map each metric into the range of $[0, 1]$. The social metric of node $v$ in $(t-\omega, t)$ is denoted as $m^u_{v,t}$, where $u$ indicates it is a normalized metric. As the tie strength is a proportion value, it does not need a mapping process to normalize it. Therefore, we have

$$TS^u_{v,t}(d) = TS_{v,t}(d) \tag{14}$$

To find the relay nodes in social DTNs, we only need to compare the values of social metrics. We can use a nonlinear monotone increasing function $f(x) = \arctan(x)$ to map the metrics into the range [0,1] without compromising the ranking of social metrics. For the betweenness, its variation range is large, with the minimum value 0 to the maximum theoretical value $(n-1)(n-2)$. If we directly use the arc tangent function to normalize it, some relatively high betweenness value may be mapped into a very small value which makes it hard to differentiate between the high betweenness values. Instead, we first divide the betweenness by the number of nodes $n$, and then normalize it as

$$B^u_{v,t} = \arctan \frac{B_{v,t}}{n} \times \frac{2}{\pi} \tag{15}$$

where $n$ is the maximum betweenness value based on the observation from real datasets. Even if there exist several values larger than $n$ in some time windows, they have no significant impact on the ranking of all nodes' betweenness after handling them by Equation (15).

The initial similarity value ranges from 0 to $n$. We first divide it by $\sqrt{n}$ and then map it as

$$S^u_{v,t}(d) = \arctan\frac{S_{v,t}(d)}{\sqrt{n}} \times \frac{2}{\pi} \tag{16}$$

The actual maximum value of similarity is about $\sqrt{n}$ in the real datasets. Similar to betweenness, some similarity values outside $\sqrt{n}$ have no effect on the ranking of all nodes' similarity after the mapping with Equation (16).

Then we can apply the improved Kalman Filter to obtain TSMs based on the normalized window metrics $B^u_{v,t}$, $S^u_{v,t}$, $TS^u_{v,t}$, respectively.

### B. TSM-Based Forwarding Mechanism

Our prediction model is general and can be applied to calculate other social metrics for typical opportunistic forwarding schemes such as SimbetTS, BubbleRap. We show the performance of some existing opportunistic forwarding schemes in our evaluation section using the TSM prediction model instead of their own social metric calculation methods.

To take full advantage of the time-varying metrics predicted with TSMs model for more efficient data forwarding, we propose a TSMs-based staged opportunistic forwarding mechanism, which can be abbreviated as TSMF. The transmission stage of $v$ is determined based on if the future time-varying *similarity* and *tie strength* $S^u_{v,t}$, $TS^u_{v,t}$ have values. When there exists a value for the *similarity*, it indicates that the node has the common neighbors with the destination. The relation between the destination and the node is closer than that with another node which has a large *betweenness centrality*. *Tie strength* shows that the node has contact with the destination, which is a direct relation. Therefore, the node with no zero *tie strength* is the closest to the destination.

In our design, we consider the noise in the recursive prediction process in Equation (10), so TSMs may have very small values while they should be zero. We add a threshold value to facilitate the correct determination of a transmission stage. Assume the number of nodes in the network is 100, if $S_{v,t}(d) = 1$, we have $S_{v,t}(d) \approx 0.06$ by Equation (16). So we could set a relative value 0.01 as the threshold $\theta$, below which the time-varying *similarity* value is considered to be zero. As we normalize the social metric values, we can apply the same threshold $\theta$ to differentiate between noise and data for all three social metrics.

Next we give the definitions as follows. If $TS_{v,t}(d)$ is larger than $\theta$, then node $v$ is $d$'s temporal neighbor (TN). If $TS_{v,t}(d)$ is less than $\theta$ while $S_{v,t}(d)$ is larger than $\theta$, $v$ is $d$'s 2-hop temporal neighbor (2-hopTN). So our staged opportunistic forwarding mechanism can be described as Algorithm 1.

This process can be summarized as follows. If the data carrier $i$ is neither the destination $d$'s TN nor its 2-hopTN, it means $i$ is far from the destination $d$ at the current time. Then the node $i$ should forward the data to the node $j$ that has the higher temporal centrality at this stage. The node $j$ will carry the data and look for $d$, $d$'s TN, or $d$'s 2-hopTN until the data is delivered to $d$ successfully.

**Algorithm 1:** Time-varying Data Forwarding Algorithm

//when node $i$ contacts with node $j$ at time $t$;
**for** *each data* $p_k$ *that node* $i$ *carries* **do**
  **if** $j == Dst(p_k)$ **then**
    node $i$ forwards $p_k$ to $j$;
  **else if** $j \in TemNeighbor(Dst(p_k))$ **then**
    **if** $TS^u_{j,t}(Dst(p_k)) > TS^u_{i,t}(Dst(p_k))$ **then**
      node $i$ forwards $p_k$ to $j$;
  **else if** $j \in 2-hopTemNeighbor(Dst(p_k))$ **then**
    **if** $S^u_{j,t}(Dst(p_k)) > S^u_{i,t}(Dst(p_k))$ **then**
      node $i$ forwards $p_k$ to $j$;
  **else**
    **if** $B^u_{j,t} > B^u_{i,t}$ **then**
      node $i$ forwards $p_k$ to $j$;

## VI. Performance Evaluation

The simulations are based on the real datasets mentioned in Section IV-B. In order to collect the relevant information of nodes in the network, the simulation time is decomposed into three phases: warm-up phase (15% of the simulation time), data transmission phase (70%), and message delivery phase without new messages generated (15%) [20].

The simulations are carried out in three parts. First, we evaluate the impact of window size on capturing the temporal properties of node social relationship. Second, we verify the advantages of our proposed forwarding strategy TSMF. Last, to demonstrate the effectiveness of our method in predicting the time varying social metrics, we applied it in some classic opportunistic routing protocols such as SimBetTS and BubbleRap.

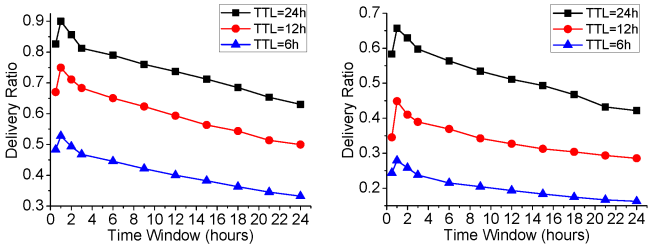
As high delivery ratio and low delivery overhead are desired while long delay are tolerable in social DTNs, we consider the following metrics [6] to evaluate the performance of forwarding strategies:

- **Data Delivery Ratio:** the ratio of the number of messages successfully transmitted to the destination to the total number of messages sent by a source node.
- **Overhead:** the distribution overhead of protocol is measured by the total number of relays used.

### A. Impact of Time Window

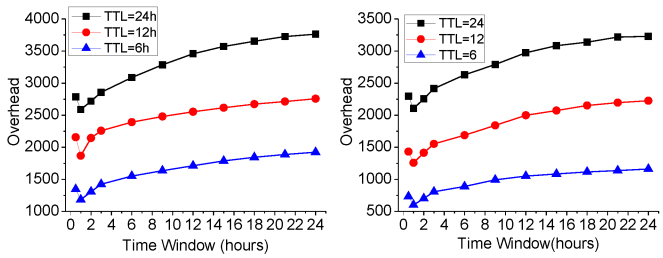
In Fig. 10 and Fig. 11, we can see that when the time window size is 1h, both datasets of Infocom06 and Cambridge have the highest data delivery ratio and the lowest overhead. With too small a window, the number of contacts is too small, so the calculations of social metrics are inaccurate.

Fig. 10 shows that when the size of time window is larger than 1h , there is a sharp reduction of the data delivery ratio initially and then the reduction speed becomes smaller as the window size further increases. When the time window size is relatively large, the variation of its value has little




(a) Infocom06 (b) Cambridge

Fig. 10: Delivery ratio of different time windows




(a) Infocom06 (b) Cambridge

Fig. 11: Overhead of different time windows

impact on the data delivery ratio. It verifies that our snapshot-based method with an appropriate time-varying window size is efficient.

Increasing the time window initially helps to more accurately calculate the time-varying social metrics, but too large a window will reduce the sensitivity of the metrics to the network dynamics.

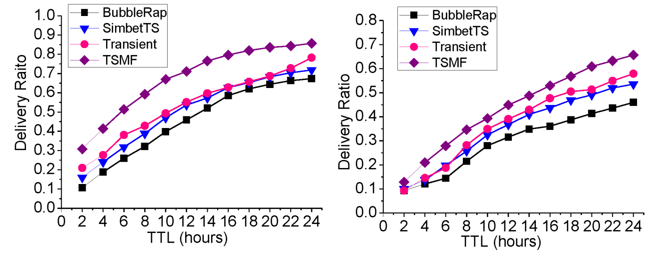
This verifies that our snapshot-based method with an appropriate time-varying window size is efficient. As mentioned in Section IV-B, the window size can be *adapted* in different network environments to better capture the social relationships and build more accurate social metrics.

### *B. Performance of Time-varying Forwarding Strategy*

We compare TSMF with popular social routing schemes BubbleRap, SimBetTS and the recent proposed Transient presented in the Section II. In Fig. 12 and Fig. 13, TSMF obtains the highest data delivery ratio and the lowest overhead. The delivery ratio is 58.2%, 33.2% and 22.1% higher than the three peer schemes, while the overhead is 38.5%, 23.7% and 25.2% lower. With more accurate prediction of TSMs, the relay nodes can be better selected in TSMF to improve the routing efficiency. Besides, with the increase of TTL (i.e., the time of data to live in social DTNs), more data could be delivered successfully with more relaxed time constraint.

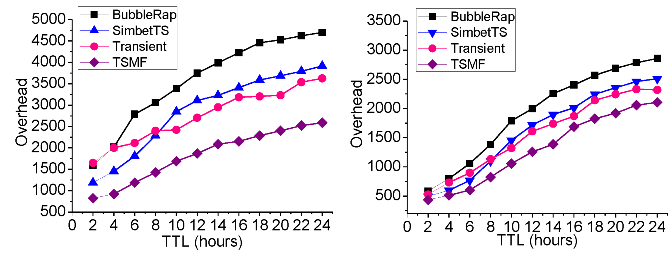
### *C. Impact of Time-varying Social Metrics*

To evaluate the advantages of our new approach in calculating time-varying social metrics, we apply it to incorporate the time-varying feature into the original metrics adopted by



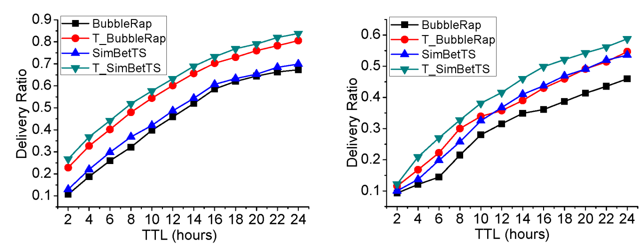

(a) Infocom06 (b) Cambridge

Fig. 12: Delivery ratio of different protocols



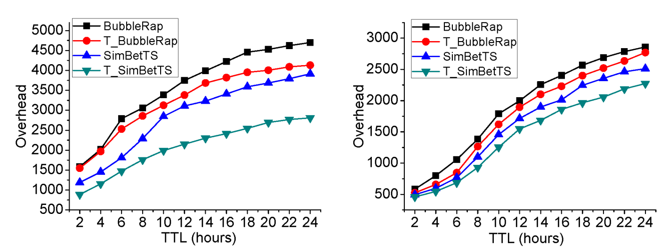

(a) Infocom06 (b) Cambridge

Fig. 13: Overhead of different protocols




(a) Infocom06 (b) Cambridge

Fig. 14: Impact of TSMs: delivery ratio




(a) Infocom06 (b) Cambridge

Fig. 15: Impact of TSMs: overhead

SimBetTS and BubbleRap, and label them as T_SimBetTS and T_BubbleRap, respectively.

In Fig. 14 and Fig. 15, T_SimBetTS and T_BubbleRap have much higher data delivery ratio and lower overhead as compared to SimBetTS and BubbleRap which use conventional static social metrics. In the dataset of Cambridge, with respect to BubbleRap, the data delivery ratio of T_BubbleRap increases on average 27.9% while its overhead decreases on average by 10.4%. With respect to SimBetTS, the data delivery ratio of T_SimBetTS increases on average by 22.6% while its overhead reduces on average by 18.7%. Fig. 14 shows that with the increase of TTL, the superiority of T_SimBetTS and T_BubbleRap is more obvious, and moreover, the overhead also reduces. The reason is that the number of hops in the delivery of message increases with a larger TTL.

This demonstrates that our proposed time-varying method for calculating the social metrics is very effective in improving the accuracy of metrics thus facilitating more efficient data delivery in social DTNs.

## VII. Conclusion

The aim of this work is to develop some general algorithms to more accurately capture time varying features of social metrics to ensure higher data forwarding performance in social DTNs. We first analyze two groups of datasets to show the limitation of traditional metrics derived from the long-term statistics. We then introduce the network model using time-varying contact graphs, based on which we observe the important temporal properties of node contacts, *durative* and *periodicity*, and derive the corresponding time-varying characteristics of the social metrics. We further exploit these properties to more accurately predict nodes' time-varying metrics, and design a real-time opportunistic forwarding mechanism. Our results from trace-driven simulations demonstrate that our time-varying social metrics help to significantly improve the performance of popular opportunistic social routing protocols, and verify that our forwarding mechanism can help further improve the data delivery.

## VIII. Acknowledgement

This work was supported in part by China 863 Program 2015AA01A705, US NSF 1526843, and China NSFC 61271185.

## References

[1] E. M. Daly and M. Haahr, Social network analysis for routing in disconnected delay-tolerant MANETs, In *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pp. 32-40, September 2007.

[2] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, Efficient Routing in Intermittently Connected Mobile Networks: The Multiple-Copy Case, *IEEE/ACM Trans. Netw.*, vol. 16, pp. 77-90, Feb. 2008.

[3] A. Lindgren, A. Doria, and O. Scheln, Probabilistic Routing in Intermittently Connected Networks, *ACM SIGMOBILE CCR* vol. 7, no. 3, pp. 19-20, 2003.

[4] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks, *In Proc. IEEE INFOCOM* vol. 6, pp. 1-11, 2006.

[5] E. M. Daly et al.Social network analysis for information flow in disconnected delay-tolerant MANETs, *IEEE Transactions on Mobile Computing*, vol. 8, pp. 606-621, May 2009.

[6] P. Hui, J. Crowcroft, and E. Yoneki, BubbleRap: Social-Based Forwarding in Delay Tolerant Networks, *IEEE Transactions on Mobile Computing*, vol. 10, no. 11, pp. 1576-1589, Nov. 2011.

[7] W. Gao, G. Cao, T. La Porta, and J. Han, On exploiting transient social contact patterns for data forwarding in delay-tolerant networks, *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 151-165, 2013.

[8] A. Casteigts, P. Flocchini, W. Quattrociocchi, N. Santoro, Time-varying graphs and dynamic networks,International Journal of Parallel, *Emergent and Distributed Systems*, vol. 27, no. 5, pp. 387-408, 2012

[9] A. M. Nguyen, N. Crespi, Social-DTN: Why Social Networking Services is More Fruitful to Mobile Delay-tolerant Networks?, In *Proceedings of International Conference on Ultra Modern Telecommunications & Workshops (ICUMT)*, 2009.

[10] Haggle project. [Online]. Available: http://www.haggleproject.org, 2004.

[11] Freeman, L.C., Centered graphs and the construction of ego networks, *Mathematical Social Sciences*, pp. 291-304, 1982.

[12] P. V. Marsden, Egocentric and sociocentric measures of network centrality, *Social networks*, vol. 24, no. 24, pp.407-422, 2002.

[13] T. Hossmann, T. Spyropoulos, and F. Legendre, Know thy neighbor: Towards optimal mapping of contacts to social graphs for dtn routing, In *Proceedings of IEEE Infocom*, pp. 1-9, 2010.

[14] M.E.J. Newman, A Measure of Betweenness Centrality Based on Random Walks, *Social Networks*, vol. 27, no. 1, pp. 39-54, Jan. 2005.

[15] Yule, G.U and Kendall, M.G. (1950), An Introduction to the Theory of Statistics, *14th edn*, page 268, 1968

[16] L.C. Freeman, Centrality in social networks: Conceptual clarification, *Social Networks*, vol. 1, no. 3, pp. 215- 239, 1979.

[17] T. Opsahl, F. Agneessens, and J. Skvoretz, Node centrality in weighted networks: Generalizing degree and shortest paths, *Social Networks*, vol. 32, no. 3, pp. 245-251, 2010.

[18] M. McPherson, L. Smith-Lovin, and J. M. Cook, Birds of a feather: Homophily in social networks, *Annual Review of Sociology*, vol. 27, no. 1, pp. 415-444, 2001.

[19] M.S. Granovetter, The strength of weak ties, *American Journal of Sociology*, vol. 78, no. 6, pp. 1360-1380.

[20] P. Costa, C. Mascolo, M. Musolesi, and G.-P. Picco, Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks, In *IEEE JSAC Special issue on Delay-Tolerant Networks*, 2008.

[21] J. Link, N. Viol, A. Goliath and K. Wehrle, SimBetAge: Utilizing temporal changes in social networks for pocket switched networks, In *Proceedings of ACM Workshop on User-provided Networking*, 2009.

[22] N. Kayastha , D. Niyato , P. Wang and E. Hossain, Applications, architectures, and protocol design issues for mobile social networks: A survey, *Proc. IEEE*, vol. 99, no. 12, pp. 2130-2158, 2011.

[23] T. Hossman, F. Legendre, From Contacts to Graphs: Pitfalls in Using Complex Network Analysis for DTN Routing, In *IEEE Infocom*, 2009.

[24] S. Bolboaca, L. Jantschi, Pearson versus Spearman, Kendall's Tau correlation analysis on structure-activity relationships of biologic active compounds, *Leonardo Journal of Sciences*, vol. 5, no. 9, pp. 179-200, 2006.

[25] J.M. Anthonisse, The Rush in a Graph, *Mathematische Centrum*, 1971.

[26] L.C. Freeman, A Set of Measures of Centrality Based on Betweenness, *Sociometry*, pp. 35-41, 1977.

[27] Conan, Vania, Jérémie, Leguay, and Timur Friedman, Characterizing pairwise inter-contact patterns in delay tolerant networks, In *Proceedings of the 1st international conference on Autonomic computing and communication systems*, 2007.

[28] Z. Zhang, Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: Overview and challenges, In *IEEE Communications Surveys and Tutorials*, vol. 8, no. 1, pp. 24-37, Jan. 2006.

[29] Picu, Andreea and Spyropoulos, Thrasyvoulos, Forecasting DTN performance under heterogeneous mobility: The case of limited replication, In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, pp. 569-577, 2012.

[30] Nelson S C, Bakht M, Kravets R. Encounter-Based Routing in DTNs, In *INFOCOM*, pp. 846-854, 2009.

[31] Hui P, Chaintreau A, Scott J, et al. Pocket Switched Networks and Human Mobility in Conference Environments, In *Acm Sigcomm Workshop on Delay-tolerant Networking*. pp. 244–251, 2005.