Characterising Temporal Distance and Reachability in Mobile and Online Social Networks

John Tang  
Computer Laboratory  
University of Cambridge  
jkt27@cam.ac.uk

Mirco Musolesi  
Computer Laboratory  
University of Cambridge  
mm753@cam.ac.uk

Cecilia Mascolo  
Computer Laboratory  
University of Cambridge  
cm542@cam.ac.uk

Vito Latora  
Dipartimento di Fisica  
University of Catania  
latora@ct.infn.it

ABSTRACT
The analysis of social and technological networks has attracted a lot of attention as social networking applications and mobile sensing devices have given us a wealth of real data. Classic studies looked at analysing static or aggregated networks, i.e., networks that do not change over time or built as the results of aggregation of information over a certain period of time. Given the soaring collections of measurements related to very large, real network traces, researchers are quickly starting to realise that connections are inherently varying over time and exhibit more dimensionality than static analysis can capture.

In this paper we propose new temporal distance metrics to quantify and compare the speed (delay) of information diffusion processes taking into account the evolution of a network from a global view. We show how these metrics are able to capture the temporal characteristics of time-varying graphs, such as delay, duration and time order of contacts (interactions), compared to the metrics used in the past on static graphs. We also characterise network reachability with the concepts of in- and out-components. Then, we generalise them with a global perspective by defining temporal connected components. As a proof of concept we apply these techniques to two classes of time-varying networks, namely connectivity of mobile devices and interactions on an online social network.

Categories and Subject Descriptors
C.2.1 [Network Architecture and Design]: Network Topology; C.2.0 [General]: Data communications

General Terms
Measurement, Algorithms, Theory

Keywords
Temporal Graphs, Temporal Metrics, Temporal Efficiency, Social Networks, Complex Networks, Information Diffusion

1. INTRODUCTION
The appearance of abundant and fine grained data about social network interactions has sparked numerous investigations into the properties of human interactions [9, 10]. What has become increasingly clear is that the time dimension of these interactions have often been neglected or understated while developing analytical methods for social and complex network analysis.

We argue that static metrics such as path length, clustering coefficient and centralitity [14], to name a few, are sufficient where temporal information is not inherent in the network but give a too coarse-grained view in networks where the temporal dynamics is an essential component of the phenomenon under observation such as human interactions over time.

Past research by Kempe et al. proposed a temporal network model with time labelled edges where paths need to obey the time order of the appearance of edges [8]. However, this model does not allow for analysis of frequency of contacts between nodes or groups, nor does it handle temporally disconnected nodes i.e., where there is no time respecting transitive path between two nodes over time. Similarly, in [11] Kostakos presented the concept of temporal graphs and an equivalent measure of delivery time between nodes of a temporal graph. However this provides a skewed indication of the global delay of the information diffusion process since it does not take into account pairs of nodes for which a transitive path does not exist. Also the lack of normalisation over nodes or time do not lend for easy comparison between networks. In [10] the authors analyse information dissemination processes focussing on identifying the diffusion of the most recent piece of information about a certain topic in a social network. We instead are interested in measuring the smallest delay path of generic information spreading processes. Spatio-temporal aspects have also been studied for the analysis of delay and data delivery in DTN networks [7, 3]. The Kempe-Kleinberg model has also been adapted for social networks analysis [5, 13, 1], however the focus of these works is on the local characteristics of time-varying networks; global aspects of the information processes in these networks are not captured.

In this paper we present new metrics related to temporal distance and reachability and evaluate how these are useful to capture properties at a fine granularity with a global and local view. The key measure that we propose is the average temporal path length of a network quantifies how fast information spreads to all nodes by means of transitive connections between them. From this measure, we derive others,
A temporal graph can be represented by means of a sequence of time windows, where for each window we consider a snapshot of the network state at that time interval. The metrics we developed over this view of the temporal graph retain the time ordering, repeated occurrences of connections between nodes, contact time and deletions of edges.

We now formally introduce the definition of temporal graph $G_t^\omega$. Given a real network trace starting at $t_{\min}$ and ending at $t_{\max}$ we define a contact between nodes $i, j$ at time $s$ with the notation $R_{ij}^s$. A temporal graph $G_t^\omega (t_{\min}, t_{\max})$ with $N$ nodes consists of a sequence of graphs $G_{t_{\min}}, G_{t_{\min}+w}, \ldots, G_{t_{\max}}$, where $w$ is the size of each window in some time unit (i.e., seconds). Then $G_t$ consist of a set of nodes $V$ and a set of edges $E$, such that $i,j \in V$, if and only if there exists $R_{ij}^s$ with $t \leq s \leq t + w$.\(^2\) We now introduce the temporal distance metric and then the global and local metrics which we have derived from this.

### 2.1 Temporal Distance

Given two nodes $i$ and $j$ we define a temporal path:

$$p_{ij}^q(t_{\min}, t_{\max})$$

(1)

to be the set of paths starting from $i$ and finishing at $j$ that pass through the nodes $n_1^t \ldots n_s^t$, where $t_{s-1} \leq t_i$ and

\(^1\)Contact in this paper expresses the general concept of a node having some sort of interaction with another node such as physical proximity or exchange of a message.

\(^2\)The limit case is a time window with duration equal to the minimum interval between the appearances of two consecutive contacts. By selecting this window size, there is no approximation in the calculation of the temporal metrics.

To compute $d_{t_{\min}}^q(t_{\min}, t_{\max})$ we have implemented a depth first search algorithm that gives the distance from a source node $i$ to all other nodes. The algorithm assumes global knowledge of the temporal graph and keeps track of two global lists, $D$ and $R$, indexed by node identifier. $D$ keeps track of the number of temporal hops to reach a node and $R$ keeps track of nodes that are reached. We initialise the value of every nodes of $D$ to 1 and $R$ to False. Starting with the first time window, we check that the source node $i$ has been sighted. If so, we perform a depth first search (DFS) to see if any unreached nodes have a path to a node that was reached in a previous window. The maximum depth of DFS is dictated by the horizon $h$ and if there are more than one path we choose the shortest. If a node $j$ is reachable then we set $R[j] = True$ otherwise we increment the distance $D[j]$.

If the source node $i$ is not reachable then we increment all $D[j]$ since we cannot establish a transitively connected path from the source. We then repeat this for the next window.

### 2.2 Example

As pointed out in the introduction, we argue that aggregated (i.e., static) graphs are unable to model temporally rich networks since they assume contact between nodes occur all at once. Let us consider the temporal graph in Fig-
ure 1 and its static version in Figure 2 where all contacts are aggregated into a single graph. If node A wanted a piece of information to reach F, according to the static graph it could do so via nodes B, C, D, and E. Also, reversing the path, if node F wanted to reach A it could do so i.e., suggesting paths are symmetric. In fact over time, contacts between A and F occur in the wrong time order to facilitate this. As we can see, the static graph incorrectly showed that information could spread between node A and node F.

We now show how our algorithm calculates the temporal distance between nodes in the network.

Starting with the first window we calculate the reachability of a message sent from node A. Figure 3 shows the snapshot of contact graph at $t = 1$ and the upper table shows the state of lists $R$ and $D$ after the initialisation phase. We first check if we can see the source node A. Since node A appears in this first window, $R[A]$ is set to True. We then iterate through every other node in the window to check for reachability. Since there is a path between A and B and also since A was reached already we update $R[B]$ to True. However for node C, there are no contacts to any other nodes so we increment the distance $D[C]$. The same applies to nodes D, E and F and the lower table shows the state of $D$ and $R$ after processing the first window. The second window is shown in Figure 4. We iterate through all unreached nodes C, D, E and F and perform DFS to see if they can be reached via already reached nodes i.e. A or B. As we can see there are contacts amongst the unreached nodes, however none are with A or B so again the distance $D$ for nodes $C$, $D$, $E$ and $F$ are incremented. The state of D and R are shown in Figure 4 after processing the second window.

In the third and final window starting from node C, we check if there is a path to a previously reached node. In this case performing DFS gives us two nodes we can reach D and B in the current window, but only node B has been reached in a previous window. We only care that there is a valid path not the number of hops within the current window, so we set $R[C] = True$. Since the value of $D[C]$ is 3 and $R[C]$ is True, we now know that a message from node A can reach node C in 3 time windows. Therefore the temporal distance $d_{AC} = 3$. For node D there is a path to node C and node B, but since only node B was reached in a previous window we use this path and set $R[D]$ to True. For nodes E and F, a message from node A has still not arrived and so the final state shown in Figure 5 reflects this. For all values of $R$ that are False we can treat the distance $D$ as $\infty$.

Table 1 shows the temporal path length calculated for every node pair, where the diagonal describes when a node was first seen by another node. As we mentioned earlier paths in static undirected graphs are assumed to be symmetric, for example in Figure 2 there is a path between node A to C and vice versa. However, in Table 1 this is not the case due to the ordering of the contacts and this can be verified visually in Figure 1.

### 2.3 Global Temporal Metrics

Global temporal metrics capture the dynamics of the whole network, in particular how easy information flows from source to destination across the whole time space. In the spirit of static global efficiency $E_{global}^{12}$, we define the temporal efficiency $E_{Ti,j}^{h}$ between nodes $i$ and $j$ and between the time interval $t_{min}$ to $t_{max}$ as:

$$E_{Ti,j}^{h}(t_{min}, t_{max}) = \frac{1}{d_{ij}^{h}(t_{min}, t_{max})}$$

where temporally disconnected nodes intuitively have $E_{Ti,j}^{h} = 1/\infty = 0$. Therefore, given a horizon $h$, we can then define the characteristic shortest temporal path length $L^{h}$ and temporal global efficiency $E_{global}^{h}$ for a temporal graph as:

$$L^{h}(t_{min}, t_{max}) = \frac{w}{N(N-1)} \sum_{ij} d_{ij}^{h}(t_{min}, t_{max})$$

$$E_{global}^{h}(t_{min}, t_{max}) = \frac{1}{N(N-1)} \sum_{ij} E_{Ti,j}^{h}(t_{min}, t_{max})$$

Notice that $L$ multiplies the average number of windows, by the window size $w$. This gives us a real-time in the chosen time units. To fully characterise a temporal graph, temporally disconnected nodes are captured in the average.
In the case of efficiency this is straightforward since temporally disconnected node pairs have a zero efficiency. In the case of temporal path length we assume that information expires after a certain time period i.e. \( t_{\max} \). Therefore, the maximum temporal length that we consider is \( (t_{\max} - t_{\min}) \).

### 2.4 Local Temporal Metrics

Local temporal metrics capture the dynamics of each node and its neighbours across the whole time space. The generalisation of the local efficiency \( E_{\text{loc}} \) for temporal graphs we propose is as follows.

We first define \( N_i(t_{\min}, t_{\max}) \) as the set of all first-hop neighbours seen by node \( i \) at least once in the time interval \([t_{\min}, t_{\max}]\). We indicate as \( k_i(t_{\min}, t_{\max}) \) the number of nodes in the set \( N_i(t_{\min}, t_{\max}) \). We then consider the sequence of subgraphs \( G^N_i(t_{\min}, t_{\max}) \), \( t = t_{\min}, t_{\min} + w, \ldots, t_{\max} \) where each \( G^N_i(t_{\min}, t_{\max}) \) is the neighbour subgraph of node \( i \), considering only the nodes in \( N_i(t_{\min}, t_{\max}) \) and retaining the edges from \( G_{t_{\min}} \).

We can define the local efficiency of node \( i \) in the time window \([t_{\min}, t_{\max}]\) as:

\[
E_{\text{loc}}(t_{\min}, t_{\max}) = E_i(t \in [t_{\min}, t_{\max}]) \tag{6}
\]

that is the efficiency of the time varying graph of the first neighbours of \( i \) in the time window \([t_{\min}, t_{\max}]\), i.e. the shortest-path for time-varying graphs are computed for \( G^N_i(t_{\min}, t_{\max}) \), \( t \in [t_{\min}, t_{\max}] \). Note that by definition, for \( E_{\text{loc}} \) the horizon is always 1 since we are only considering the direct neighbours of node \( i \).

### 2.5 Temporal Components

As discussed above, a node \( j \) is reachable in the time interval \([t_{\min}, t_{\max}]\) from node \( i \) if there is a temporal path from node \( i \) to node \( j \) or, in other words, if a message can be delivered from node \( i \) to node \( j \) in that time interval.

In static analysis, individual node reachability is defined by the in-component and out-component, which define the set of nodes that can reach and be reached by a node \( i \); and collective reachability among groups of nodes in a network is defined by the connected components \( CC \), which defines the sets of nodes that can reach each other such that there may be disjoint islands of nodes.

Formally, for an undirected static graph \( G = (V, E) \) this is defined as the maximal set of vertices \( C \subseteq V \) such that for every pair of vertices \( i, j \in C \), there exists a path from \( i \) to \( j \) \cite{2}. This definition means that each node can only belong to a single component. In the static graph (Figure 2), since it assumes all nodes have a path to all other nodes, there is only one component set consisting of nodes \( \{A, B, C, D, E, F\} \). Such a graph with a single connected component is also described as connected.

In a directed static graph, reachability can be defined in terms of weakly connected components or strongly connected components. In the latter case, the strongly connected component for a directed graph \( G = (V, E) \) is the maximal set of vertices \( C \subseteq V \) such that for every pair of vertices \( i, j \in C \), there exists both a path from \( i \) to \( j \) and from \( j \) to \( i \) \cite{2}. In the former case the direction of links are ignored.

We now extend these concepts to our temporal model with the aid of example temporal graph \( II \) (Figure 6), noting that the resulting aggregated static graph equal to the previous example (Figure 2). The calculated temporal path length matrix is shown in Table 2.

![Figure 6: Example Temporal Graph II, \( G_t(0, 3) \), \( h = 2 \) and \( w = 1 \).](image)

![Figure 7: Temporally connected components of temporal graph II (Figure 6).](image)

To define the temporal out-component \( OUT^h_i(t_{\min}, t_{\max}) \) of a node \( i \) as the set of nodes that \( i \) can reach in the time interval \([t_{\min}, t_{\max}]\) with horizon \( h \). Analogously, we define the temporal in-component \( IN^h_i(t_{\min}, t_{\max}) \) of a node \( i \) as the set of the nodes from which node \( i \) can be reached in the time interval \([t_{\min}, t_{\max}]\) with horizon \( h \).

<table>
<thead>
<tr>
<th>( A )</th>
<th>( B )</th>
<th>( C )</th>
<th>( D )</th>
<th>( E )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Temporal path length matrix for temporal graph II (Figure 6).

\[ ^3 \text{Though note that a weakly temporal connected component can be shown to always be equal to the static connected component of the graph which renders its definition trivial.} \]
connected components can be overlapping. Indeed, the transitivity properties are different in temporal graphs since we have temporal ordering and, therefore, the property of reachability is not symmetric. For example, in the temporal graph in Figure 6 the sets of nodes \((A, B, C, D)\) can communicate to each other in the time interval \([t_1, t_3]\). The same happens for the nodes of the set \((C, E, F)\). In other words, we have two distinct temporally connected components as shown in Figure 7. However, the fact that the node \(C\) belongs to both components does not imply that the two (sub-)components form a single component given by the union of the two as in the static case given the asymmetric temporal reachability property.

Through this simple example, we have shown that since static analysis ignores the time ordering it cannot capture the true connected components. In more practical terms, this means that static analysis gives us a misleading estimation of the connectedness of the network. This aspect is essential for the analysis of real-world phenomena, such as message dissemination and epidemics spreading.

3. EVALUATION

In our evaluation we use three datasets: Bluetooth traces of people at the 2005 INFOCOM conference [6], campus Bluetooth traces of students and staff at MIT [4] and interactions between a large group of members of a large online social network, namely Facebook users affiliated with the London network [15]. We shall refer to these as INFOCOM, REALITY and FACEBOOK, respectively. Table 3 describes the characteristics of each set of traces.

The INFOCOM traces were collected in a conference environment using Bluetooth colocation scanning every 2 minutes. With 41 nodes it is quite a small trace but temporally dense in that there are a high number of contacts per day. The REALITY traces were collected at the MIT campus between Bluetooth phones sightings of students, research staff and professors, with Bluetooth scanning every 5 minutes. We split these two traces into individual days. The FACEBOOK traces were crawled over a one year period (March 2007 to February 2008) from the members of the London network. We consider two types of user interactions, the posting of contents on a user webpage (called wall postings in Facebook) and comments on user photos. Pairs of nodes with less than 10 interactions were filtered so that only the most active users remain. To make the experimental results comparable between the INFOCOM and REALITY traces we fix the window size \(w\) to 5 minutes which is equal to the longest Bluetooth scanning rate of the REALITY trace. We discuss the effects of different values of window size in Section 3.4. Given the different time scale and the fact that the FACEBOOK traces are very sparse, we use a window size of 1 hour.

Table 3: Experimental Data Sets.

<table>
<thead>
<tr>
<th>Start</th>
<th>INFOCOM</th>
<th>REALITY</th>
<th>FACEBOOK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>4 days</td>
<td>280 days</td>
<td>12 Months</td>
</tr>
<tr>
<td>Times</td>
<td>day1:6pm-12pm</td>
<td>12am-12pm</td>
<td>day1:6pm-12pm</td>
</tr>
<tr>
<td></td>
<td>day2:12am-12pm</td>
<td>12am-12pm</td>
<td>day2:12am-12pm</td>
</tr>
<tr>
<td></td>
<td>day3:12am-12pm</td>
<td>12am-12pm</td>
<td>day3:12am-12pm</td>
</tr>
<tr>
<td></td>
<td>day4:12am-5pm</td>
<td>12am-12pm</td>
<td>day4:12am-5pm</td>
</tr>
<tr>
<td>No. of nodes</td>
<td>41</td>
<td>100</td>
<td>164,135</td>
</tr>
<tr>
<td>Contacts(avy. per day)</td>
<td>4817</td>
<td>231</td>
<td>17835</td>
</tr>
<tr>
<td>Granularity</td>
<td>120 secs.</td>
<td>300 secs.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

3.1 Comparison with Static Metrics

Firstly, as a comparison between the temporal and the static metrics, we show the results calculated for the INFOCOM data set. As argued before, paths in static graphs ignore duration of contacts, inter-contact time, recurrent contacts and time ordering of contacts and so overestimate the number of connected node pairs and underestimate the path lengths. Table 4 shows calculations for both static and temporal path lengths, \(L\). As a note, since our temporal \(L\) metric presented in Equation 4 is in real time, it is hard to compare with static \(L\). To bridge the gap we show temporal \(L^*\) which is calculated as the average shortest node to node hop that obeys time ordering of edges. This is fair since temporal \(L\) uses the same time ordered path but measured in terms of elapsed time. As we can see in the static results for Day 1, path length is low. Now looking at the temporal aspects, we have calculated the same metrics but obeying time ordering, duration and recurrence of contacts. The third column, Disc, shows the ratio of disconnected node pairs. In the case of static graphs, there were no disconnected node pairs. As we can see temporal \(L^* \gg\) static \(L\) and there are also much more disconnected node pairs due to the observed asymmetry and time ordering of paths. In other words, temporal \(L\) give us a better understanding of the network with respect to the temporal dimension since they can provide us an accurate measure of the delay of the information diffusion process that is not possible with traditional static metrics.

3.1.1 Distance Metrics

With respect to the temporal reachability of the network, the importance of considering the temporal dimension is apparent. In fact, if we compare the static and temporal connected components (CC) for INFOCOM (Table 4) we can see that the static model overestimates the connectedness of the network, since it ignores time order and, therefore, it overestimates the available paths. Notice also for the first and last days, there are more temporal connected components since the days were shorter. This phenomenon is not captured in static analysis.

Let us consider for example the distributions of temporal in-component and out-component for INFOCOM days 1 and 2 (Figure 8). We can observe that the out-component is similarly large for many nodes which tells us that nodes were able to send and deliver messages easily. We would expect a similar distribution for the in-component, and indeed for the day 2 this is the case, however notice with the first, there is a drop in the number of nodes which can receive messages. This means that although the set of nodes which can successfully send and deliver messages is large, the set of nodes which actually receive these messages is smaller for the first and last days. This drop with in-component was also observed for day 4.

<table>
<thead>
<tr>
<th>Day</th>
<th>Static</th>
<th>Temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N)</td>
<td>(L)</td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>1.291</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>1.269</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>1.420</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>1.444</td>
</tr>
</tbody>
</table>

Table 4: INFOCOM Static and Temporal Metrics \((h = \text{max}, t_{\text{min}} = 12am, t_{\text{max}} = 12pm, w = 5min)\).
3.2 Temporal Efficiency of Human Contacts

We now calculate temporal $L$ from Equation 4 as a real time along with the temporal efficiency $E$. The left hand side of Tables 5, 6 and 7 show the temporal metrics for INFOCOM, REALITY and FACEBOOK, respectively. The right hand side of the tables will be discussed in the next section.

First looking at INFOCOM, recall in Table 4 that static $L$ and $L^*$ only told us the average number of hops in a path but gave us no indication of how long each hop took. Our temporal metrics give us a value that takes account of time and also captures disconnected nodes. From Table 5 we can see $L$ for Day 1: if two people started gossiping at the start of the day, it would take 19 hours to spread to all participants. We also see it is quicker to spread information in the second, third and final day of the conference at about 10 hours. From Table 3 this makes sense since on the first day participants did not start until 6pm (i.e., there is an initial delay equal to 18 hours).

What we see from the low values of $E_{glob}$ and $E_{loc}$ are that contacts between all participants. Contacts between acquaintances did not allow for a high capacity to spread information. Since temporal local efficiency $E_{loc}$ measures how people you meet interact amongst themselves, we can study this phenomena in more detail and analyse on a local view if the interaction between such acquaintances are any better for spreading information. In this case, $E_{loc}$ for each day is similar to but slightly lower than $E_{glob}$: this tells us that acquaintances do not congregate together very often.

The REALITY data set has many more days so it gives us a better overview of day to day trends. In Table 6 we show 10 consecutive Wednesdays starting from the first day of the Fall ’04 semester (8th Sep to 9th Dec 2004)\(^4\). For the first day we can see that it is slow for information to spread since $L = 23$ hours. Since both local and global efficiency are at zero, participants infrequently interacted with eachother. This makes sense since relationships are unlikely to have formed and so there are less contacts.

\(^4\)http://web.mit.edu/registrar/www/calendar0405.html

Table 5: INFOCOM Temporal Metrics ($h = 1, t_{min} = 12am, t_{max} = 12pm, w = 5min$), (shuff: fladdruns = 50).

During the subsequent Wednesdays the information spreading process is quicker and there is also a steady decrease in the average temporal path length. However still compared to the conference environment, on a campus it is twice as slow for information to spread.

The final FACEBOOK dataset gives us an indication of how information can be spread in a large online social network. Table 7 shows temporal metrics calculated for four months. We observe that the temporal path length for the first month of March is 19 days. This seems slow, but we should put this value into perspective: firstly interactions occur instantaneously since online interactions do not necessitate users to be together for extended periods of time, unlike human contacts in INFOCOM and REALITY; secondly, users generally do not reply to wall posting immediately or even at all for photo comments and so introduces natural delay between interactions. Yet for information to disseminate between all $10^4$ node pairs it takes only 19 days on average which is reflected in the low temporal efficiency values. Data diffusion is then quicker in subsequent months, taking just over half a month for all nodes to send messages to each other.

3.3 Effects of Cyclic Social Behaviour

As a null model, we compare the real data sets $G_t$ with their randomised counterpart where we have randomly reshuffled the time windows $G_t \in G_t$, destroying any inherent time order. The right hand side of Tables 5, 6 and 7 show the metrics calculated on reshuffled temporal graphs for INFOCOM, REALITY and FACEBOOK, respectively. As we can see in the two human mobility based traces INFOCOM and REALITY, in the shuffled network gives a quicker data diffusion time and higher efficiency. The reason for this is down to the cyclic behaviour of humans contacts. Humans as a collective congregate during the working hours and are more sociable during mid week. This means that there is a denser number of contacts at certain times which limits the opportunity for transitive meetings between friends to certain times of the day and decreases the speed of data diffusion. Reshuffling leads to the introduction of heterogeneity of contacts throughout a time period and introduces more opportunity for contacts throughout the day. In other words, if the distribution of delivery times is concentrated midday then shuffling spreads the concentration out throughout the day.

As far as the FACEBOOK traces are concerned, the values of the metrics for the shuffled traces are relatively close to the unshuffled ones. This tells us that the natural time ordering of user activity is organised in a way that is very effective for data diffusion, since the opportunities for data dissemination are evenly spread during the day.
to the temporal path length to capture node importance in the form of a temporal centrality measure, and to see how the maximum diffusion range evolves over time by introducing the concept of a temporal diameter.

Acknowledgments
We thank our colleagues in the NETOS Group for their comments, Ben Y. Zhao for supplying the Facebook traces, Salvatore Scellato for his help in processing them and EPSRC support through grants EP/D077273, EP/C544773 and EP/F013442.

5. REFERENCES

Table 6: REALITY Temporal Metrics 10 days (h = 1, tmin = 12am, tmax = 12pm, w = 5min),(shuffled runs = 50).

<table>
<thead>
<tr>
<th>Date</th>
<th>L</th>
<th>E_glob</th>
<th>E_loc</th>
<th>L</th>
<th>E_glob</th>
<th>E_loc</th>
</tr>
</thead>
<tbody>
<tr>
<td>08 Sep</td>
<td>23h 15m</td>
<td>0.000</td>
<td>0.000</td>
<td>21h 58m</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>15 Sep</td>
<td>22h 47m</td>
<td>0.001</td>
<td>0.000</td>
<td>19h 55m</td>
<td>0.024</td>
<td>0.007</td>
</tr>
<tr>
<td>22 Sep</td>
<td>22h 53m</td>
<td>0.001</td>
<td>0.000</td>
<td>20h 42m</td>
<td>0.019</td>
<td>0.007</td>
</tr>
<tr>
<td>29 Sep</td>
<td>22h 20m</td>
<td>0.001</td>
<td>0.000</td>
<td>17h 44m</td>
<td>0.037</td>
<td>0.009</td>
</tr>
<tr>
<td>06 Oct</td>
<td>22h 14m</td>
<td>0.001</td>
<td>0.000</td>
<td>16h 23m</td>
<td>0.041</td>
<td>0.011</td>
</tr>
<tr>
<td>13 Oct</td>
<td>21h 37m</td>
<td>0.004</td>
<td>0.000</td>
<td>14h 57m</td>
<td>0.055</td>
<td>0.013</td>
</tr>
<tr>
<td>20 Oct</td>
<td>21h 45m</td>
<td>0.003</td>
<td>0.000</td>
<td>17h 4m</td>
<td>0.031</td>
<td>0.007</td>
</tr>
<tr>
<td>27 Oct</td>
<td>22h 1m</td>
<td>0.001</td>
<td>0.002</td>
<td>15h 19m</td>
<td>0.050</td>
<td>0.013</td>
</tr>
<tr>
<td>03 Nov</td>
<td>21h 6m</td>
<td>0.004</td>
<td>0.000</td>
<td>16h 17m</td>
<td>0.043</td>
<td>0.012</td>
</tr>
<tr>
<td>10 Nov</td>
<td>20h 5m</td>
<td>0.004</td>
<td>0.000</td>
<td>14h 25m</td>
<td>0.061</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Table 7: FACEBOOK Temporal Metrics 4 months (h = 1, tmin = 18th of each month, tmax = 18th of following month, w = 1hour),(shuffled runs = 1000).

<table>
<thead>
<tr>
<th>Month</th>
<th>L</th>
<th>E_glob</th>
<th>E_loc</th>
<th>L</th>
<th>E_glob</th>
<th>E_loc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar</td>
<td>19h 08</td>
<td>1.83E-04</td>
<td>1.89E-06</td>
<td>18h 19</td>
<td>2.10E-04</td>
<td>2.31E-06</td>
</tr>
<tr>
<td>Jun</td>
<td>15h 20</td>
<td>4.75E-03</td>
<td>3.23E-06</td>
<td>15h 1</td>
<td>5.70E-03</td>
<td>3.86E-06</td>
</tr>
<tr>
<td>Sep</td>
<td>17h 06</td>
<td>5.95E-03</td>
<td>2.35E-06</td>
<td>17h 7</td>
<td>7.40E-03</td>
<td>2.86E-06</td>
</tr>
<tr>
<td>Dec</td>
<td>16h 19</td>
<td>4.70E-03</td>
<td>2.49E-06</td>
<td>16h 11</td>
<td>5.90E-05</td>
<td>2.79E-06</td>
</tr>
</tbody>
</table>