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 centrality [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, namely, the temporal network efficiency (a static definition of which is contained in [12]) and connected components.

Previous work on small world effects such as the analysis based on short path length and high clustering on static graphs obtained by aggregating all the links over a certain period of time indicated that these networks are good for data diffusion due to a few edges acting as shortcuts, connecting distant nodes together [14]. However, we show that since static graphs treat all links as appearing at the same time, they do not capture key temporal characteristics such as duration of contacts[1], inter-contact time, recurrent contacts and time order of contacts along a path. For this reason, they give us an overestimate of the potential paths connecting pairs of nodes and they cannot provide any information about the delay associated with the information spreading process.

We show that our metrics are able to quantify and compare in a compact way these temporal characteristics for the study of information diffusion processes. As a proof of concept, we apply temporal metrics on conference, campus and online social network (OSN) traces and we show that the network of OSN user interactions exhibit different and slower efficiency characteristics for data diffusion than that of human contacts.

The rest of this paper is organised as follows. In Section 2 we present a formal definition of our model of temporal graphs and temporal metrics. In Section 3 we present preliminary results of the calculation of the metrics on the datasets before concluding with a discussion of the results and future work in Section 4.

2. TEMPORAL METRICS

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 $\mathcal{G}_t^w$. 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 $\mathcal{G}_t^w(t_{min}, t_{max})$ with N nodes consists of a sequence of graphs $G_{t_{min}}$, $G_{t_{min}+w}$, ..., $G_{t_{max}}$, 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:

$$p_{ij}^h(t_{min}, t_{max}) \quad (1)$$

to be the set of paths starting from i and finishing at j that pass through the nodes $n_1^{t_1} \dots n_i^{t_i}$, where $t_{i-1} \leq t_i$ and $t_{min} \leq t_i \leq t_{max}$ is the time window that node n is visited and h is the max hops within the same window t. There may be more than one shortest path. Given two nodes i and j we define the *shortest temporal distance*:

$$d_{ij}^h(t_{min}, t_{max}) \quad (2)$$

to be the shortest temporal path length. Starting from time t_{min}, this can be thought of as the number of time windows (or *temporal hops*) it takes for information to spread from a node i to node j. The *horizon h* indicates the maximum number of nodes within each window G_T which information can be exchanged, or in practical terms, the speed that a message travels. In the case of *temporally disconnected node pairs* q, p i.e., information from q never reaches p, then we set the temporal distance $d_{pq} = \infty$.

To compute $d_{ij}^h(t_{min}, t_{max})$ 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-

[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.

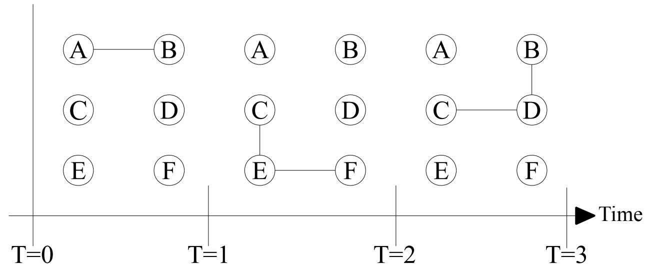


Figure 1: Example Temporal Graph I, $\mathcal{G}_t(0,3)$,$h = 2$ and $w = 1$.

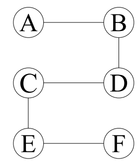
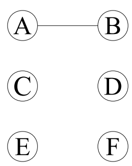


Figure 2: Example static graph based on temporal graph I (Figure 1).

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.



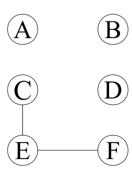
Initialization

	A	B	C	D	E	F
D	1	1	1	1	1	1
R	F	F	F	F	F	F

Variables

	A	B	C	D	E	F
D	1	1	2	2	2	2
R	T	T	F	F	F	F

Figure 3: Distance and Reachability of Window 1.

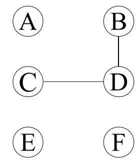


Variables

	A	B	C	D	E	F
D	1	1	3	3	3	3
R	T	T	F	F	F	F

Figure 4: Distance and Reachability of Window 2.

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 ∞.



Final

	A	B	C	D	E	F
D	1	1	3	3	4	4
R	T	T	T	T	F	F

Figure 5: Distance and Reachability of Window 3.

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.

	A	B	C	D	E	F
A	1	1	3	3	∞	∞
B	1	1	3	3	∞	∞
C	∞	3	2	3	2	2
D	∞	3	3	3	∞	∞
E	∞	3	2	3	2	2
F	∞	3	2	3	2	2

Table 1: Temporal path length for all nodes.

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_{glob} [12], we define the *temporal efficiency* $E_{T_{ij}}$ between nodes i and j and between the time interval t_{min} to t_{max} as:

$$E^h_{T_{ij}}(t_{min}, t_{max}) = \frac{1}{d^h_{ij}(t_{min}, t_{max})} \quad (3)$$

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

$$L^h(t_{min}, t_{max}) = \frac{w}{N(N-1)} \sum_{ij} d^h_{ij}(t_{min}, t_{max}) \quad (4)$$

$$E^h_{glob}(t_{min}, t_{max}) = \frac{1}{N(N-1)} \sum_{ij} E^h_{T_{ij}}(t_{min}, t_{max}) \quad (5)$$

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 disconnnected 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_{loc} for temporal graphs we propose is as follows.

We first define $\mathcal{N}_i(t_{min}, t_{max})$ 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 $\mathcal{N}_i(t_{min}, t_{max})$. We then consider the sequence of subgraphs $G_t^{\mathcal{N}_i(t_{min},t_{max})}$, $t = t_{min}, t_{min+w}, \ldots, t_{max}$ where each $G^{\mathcal{N}_i(t_{min},t_{max})}$ is the neighbour subgraph of node i, considering only the nodes in $\mathcal{N}_i(t_{min}, t_{max})$ 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:

$$E_{loc_i}(t_{min}, t_{max}) = E_T\{G_t^{\mathcal{N}_i(t_{min},t_{max})} \quad t \in [t_{min}, t_{max}]\} \quad (6)$$

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 $G_t^{\mathcal{N}_i(t_{min},t_{max})}$, $t \in [t_{min}, t_{max}]$. Note that by definition, for E_{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 eachother 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 [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 [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.

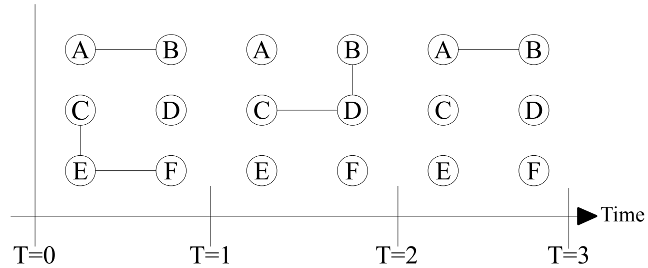


Figure 6: Example Temporal Graph II, $\mathcal{G}_t(0, 3)$, $h = 2$ and $w = 1$.

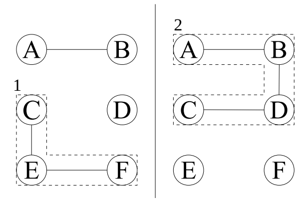


Figure 7: Temporally connected components of temporal graph II (Figure 6).

We define the *temporal out-component* $OUT_{t_i}^h(t_{min}, t_{max})$ 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* $IN_{t_i}^h(t_{min}, t_{max})$ 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.

	A	B	C	D	E	F
A	1	1	2	2	∞	∞
B	1	1	2	2	∞	∞
C	3	2	1	2	1	1
D	3	2	2	2	∞	∞
E	3	2	1	2	1	1
F	3	2	1	2	1	1

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

To define the *temporal connected components* $CC^h(t_{min}, t_{max})$, first recall that there are inherently asymmetric temporal paths between pairs of nodes and so again we can define both *weak* and *strong* temporally connected components. However, the inherent directionality of time cannot be ignored. For this reason we are only interested in providing a definition for the latter case[3]. We will then omit the word *strongly* from now on. In a temporal network we define a *temporally connected component* as the set of nodes where a temporal path exists between each member of the set or, in other words, a set where each element of the set is in the temporal out-component of each other. If there exists one single temporally connected component then we refer to the temporal graph as *temporally connected*.

In static components, an element can belong only to one component. This is not the case in temporal graphs where

[3]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.

	INFOCOM	REALITY	FACEBOOK
Start	2005-03-13	2004-07-26	2007-03-18
Duration	4 days	280 days	12 Months
Times	day1:6pm-12pm day2:12am-12pm day3:12am-12pm day4:12am-5pm	12am-12pm	12am-12pm
No. of nodes	41	100	164,135
Contacts(av. per day)	4817	231	17835
Granularity	120 secs.	300 secs.	N/A

Table 3: Experimental Data Sets.

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.

			Static		Temporal		
Day	N	$\langle k \rangle$	L	CC	L*	CC	Disc
1	37	25.7	1.291	1	4.090	5	0.28
2	39	28.3	1.269	1	4.556	2	0.13
3	38	22.3	1.420	1	4.003	2	0.19
4	39	21.4	1.444	1	4.705	3	0.14

Table 4: *INFOCOM* Static and Temporal Metrics ($h = max, t_{min} = 12am, t_{max} = 12pm, w = 5min$).

3.1 Comparison with Static Metrics

3.1.1 Distance 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 length, 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.

3.1.2 Connected Components

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 recieve messages. This means that although the set of nodes which can succesfully 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.

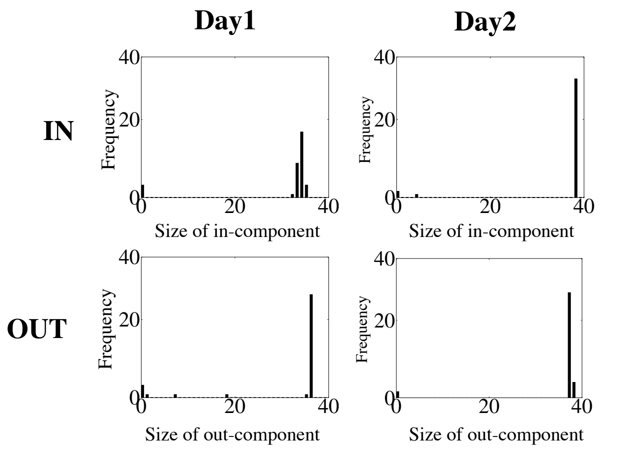


Figure 8: Frequency distributions of temporal in-component and out-component for each day of *INFOCOM*.

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

Temporal Metrics				Reshuffled		
Date	L	E_{glob}	E_{loc}	L	E_{glob}	E_{loc}
1	19h 39m	0.003	0.003	5h 29m	0.100	0.077
2	9h 6m	0.024	0.020	2h 45m	0.239	0.194
3	10h 32m	0.018	0.013	4h 6m	0.167	0.144
4	9h 55m	0.013	0.009	3h 3m	0.165	0.104

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

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 $\mathcal{G}_t$ with their randomised counterpart where we have randomly reshuffled the time windows $G_T \in \mathcal{G}_t$, destroying any inherent time order. The right hand half 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.

Temporal Metrics				Reshuffled		
Date	L	E_{glob}	E_{loc}	L	E_{glob}	E_{loc}
08 Sep	23h 15m	0.000	0.000	21h 58m	0.010	0.003
15 Sep	22h 47m	0.001	0.000	19h 55m	0.024	0.007
22 Sep	22h 53m	0.001	0.000	20h 42m	0.019	0.007
29 Sep	22h 20m	0.001	0.001	17h 44m	0.037	0.009
06 Oct	22h 14m	0.001	0.000	16h 23m	0.041	0.011
13 Oct	21h 37m	0.004	0.000	14h 57m	0.055	0.013
20 Oct	21h 45m	0.003	0.001	17h 4m	0.031	0.007
27 Oct	22h 1m	0.001	0.002	15h 19m	0.050	0.013
03 Nov	21h 6m	0.004	0.001	16h 17m	0.043	0.012
10 Nov	20h 5m	0.004	0.000	14h 25m	0.061	0.015

Table 6: *REALITY Temporal Metrics 10 days* ($h = 1, t_{min} = 12am, t_{max} = 12pm, w = 5min$)*,*($shuffledruns = 50$)*.*

Temporal Metrics				Reshuffled		
Month	L	E_{glob}	E_{loc}	L	E_{glob}	E_{loc}
Mar	19d 0h	1.83E-04	1.89E-06	18d 19h	2.10E-04	2.31E-06
Jun	15d 20h	4.70E-05	3.25E-06	15d 1h	5.70E-05	3.38E-06
Sep	17d 6h	5.80E-05	2.83E-06	17d 7h	7.40E-05	2.86E-06
Dec	16d 19h	4.70E-05	2.49E-06	16d 11h	5.90E-05	2.79E-06

Table 7: *FACEBOOK Temporal Metrics 4 months* ($h = 1, t_{min} =$ *18th of each month,* $t_{max} =$*18th of following month,* $w = 1hour$)*,*($shuffledruns = 1000$)*.*

3.4 Effects of Varying Window Sizes

We analyse how varying the window size affects the temporal metrics. By considering a larger window size the accuracy of the measurements decreases since by neglecting the order of edge appearances, the temporal path length is under-estimated as it considers links that cannot be exploited in reality. This is coupled with the higher granularity of the measurement units leading to a lower precision in the estimation of the temporal path length (which is over-estimated). However, the latter phenomenon is predominant in the traces taken into consideration, therefore we observe a higher temporal path length as window size increases. On the other hand, since it is inversely proportional to the temporal path length, temporal efficiency decreases as the window size increases.

4. CONCLUSIONS

We have presented a set of new temporal distance metrics and applied them to characterise the dynamics and data diffusion efficiency of social networks. We have introduced the concept of reachability of a network in a quantitative way by defining temporal in and out-components and connected components.

As a preliminary case study, we have provided comparable, quantitative results using three social network datasets, however the application of the temporal model and metrics can be extended to many fields that involve the study of highly dynamic networks including sensor networks, internet routing, mobility models and delay tolerant networks. Applications outside computer science include sociology and epidemiology (e.g., study of information and epidemics spreading in mobile and social networks [9]).

In this paper we have not shown the effects of increasing the horizon variable, but initial results show intuitively that the temporal path length decreases and global efficiency increase as the reach increases. There are also clear extensions 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 using 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

[1] A. Clauset and N. Eagle. Persistence and Periodicity in a Dynamic Proximity Network. In *Proc. of DIMACS Workshop on Computational Methods for Dynamic Interaction Network*, 2007.

[2] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms, Second Edition.* The MIT Press, 2nd edition, Sept. 2001.

[3] E. Daly and M. Haahr. Social network analysis for information flow in disconnected Delay-Tolerant MANETs. *IEEE Trans. Mob. Comp.*, 8(5):606–621, 2009.

[4] N. Eagle and A. Pentland. Reality Mining: Sensing Complex Social Systems. *Personal Ubiquitous Comput.*, 10(4):255–268, 2006.

[5] P. Holme. Network Reachability of Real-world Contact Sequences. *Physical Review E*, 71(4):046119–8, Apr. 2005.

[6] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket Switched Networks and Human Mobility in Conference Environments. In *Proc. of ACM SIGCOMM WDTN '05*, pages 244–251, 2005.

[7] S. Jain, K. Fall, and R. Patra. Routing in a Delay Tolerant Network. In *Proc. of ACM SIGCOMM '04*, pages 145–158, 2004.

[8] D. Kempe, J. Kleinberg, and A. Kumar. Connectivity and Inference Problems for Temporal Networks. *J. Comp. Sys. Sci.*, 64(4):820–842, June 2002.

[9] J. Kleinberg. The Convergence of Social and Technological Networks. *Commun. ACM*, 51(11):66–72, 2008.

[10] G. Kossinets, J. Kleinberg, and D. Watts. The Structure of Information Pathways in a Social Communication Network. In *Proc. of ACM SIGKDD '08*, pages 435–443, 2008.

[11] V. Kostakos. Temporal Graphs. *Physica A*, 388(6):1007–1023, Mar. 2009.

[12] V. Latora and M. Marchiori. Efficient Behavior of Small-World Networks. *Physical Review Letters*, 87(19):198701, Oct. 2001.

[13] A. Mtibaa, A. Chaintreau, J. LeBrun, E. Oliver, A. K. Pietilainen, and C. Diot. Are you Moved by your Social Network Application? In *Proc. of ACM SIGCOMM WOSN '08*, pages 67–72, 2008.

[14] D. J. Watts and S. H. Strogatz. Collective Dynamics of 'Small-world' Networks. *Nature*, 393(6684):440–2, June 1998.

[15] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. pages 205–218, Nuremberg, Germany, 2009. ACM.