Wireless Epidemic Spread in Dynamic Human Networks

Eiko Yoneki, Pan Hui and Jon Crowcroft

University of Cambridge Computer Laboratory
Cambridge CB3 0FD, United Kingdom
{firstname.lastname}@cl.cam.ac.uk

Abstract. The emergence of Delay Tolerant Networks (DTNs) has culminated in a new generation of wireless networking. New communication paradigms, which use dynamic interconnectedness as people encounter each other opportunistically, lead towards a world where digital traffic flows more easily. We focus on human-to-human communication in environments that exhibit the characteristics of social networks. This paper describes our study of information flow during epidemic spread in such dynamic human networks, a topic which shares many issues with network-based epidemiology. We explore hub nodes extracted from real world connectivity traces and show their influence on the epidemic to demonstrate the characteristics of information propagation.

1   Introduction

Increasing numbers of mobile computing devices form dynamic networks in daily life. In such environments, the nodes (i.e. laptops, PDAs, smart phones) are sparsely distributed and form a network that is often partitioned due to geographical separation or node movement. We envision new communication paradigms, using dynamic interconnectedness between people and urban infrastructure, leading towards a world where digital traffic flows in small leaps as people pass each other [15]. Delay Tolerant Networks (DTNs) [9] are a new communication paradigm to support such network environments, and our focus is a type of DTN that provides intermittent communication for humans carrying mobile devices: the Pocket Switched Network (PSN) [2].

Efficient forwarding algorithms for such networks are emerging, mainly based on epidemic protocols where messages are simply flooded when a node encounters another node. Epidemic information diffusion is highly robust against disconnection, mobility and node failures, and it is simple, decentralised and fast. However, careful tuning to achieve reliability and minimise network load is essential. Traditional naïve multiple-copy-multiple-hop flooding schemes have been empirically shown to work well in dense environments, and they provide fair performance in sparse settings – such as city-wide communications – in terms of delivery ratio and delay [2]. However, in terms of delivery cost, the naïve approach is far from satisfactory, because it creates a large amount of unwanted traffic as a side-effect of the delivery scheme. To reduce the overhead of epidemic routing, various approaches have been reported, ranging from count-,

timer- or history-based controlled flooding to location-based strategies (see Section 7 for further details).

We have previously reported an approach that uses a logical connection topology, and that uncovers hidden stable network structures, such as social networks [14] [34], from the human connectivity traces. In PSNs, social networks could map to computer networks since people carry the computer devices. We have shown improved performance by applying these extracted social contexts to a controlled epidemic strategy [13]. During this work, we have realised that further understanding of network models is essential, because the properties of human contact networks – such as community and weight of interactions – are important aspects of epidemic spread. Recently, online-based social networks have been studied; however, understanding network structures and models hidden in pervasive dynamic human networks is a still-untouched research area.

Networks represent flows of information and make it possible to characterise the complex systems of our world. A network is a map of interactions, because communication is fundamental in our society. These networks are often neither regular lattices, nor are all units connected randomly, but the interaction patterns are complex. This paper shows a preliminary study of patterns of information flow during epidemic spread in complex dynamic human networks, which share many issues with network-based epidemiology. Many studies have been conducted, and these are based either on simulation or a small collection of data. Our study uses real world data, and we believe that it gives an interesting insight on real human interactions. We consider a model for time paths based on graph evolution, called *Time-Dependent Networks*, in which links between nodes depend on a time window. We explore epidemic change by exploiting device connectivity traces from the real world and demonstrate the characteristics of information propagation. We describe preliminary empirical results, but further mathematical modelling work is outside the scope of this paper.

The rest of this paper is structured as follows. We introduce the experimental data sets in Section 2, and then describe the complexity of real world connectivity data in Section 3. We discuss the result of the epidemic spread experiments in Section 4, and the influence of hub nodes for the epidemic spread in Section 5. We describe a summary of community detection in Section 6, which is followed by the related work. Finally, we conclude the paper with a brief discussion.

2   Real World Human Connectivity Traces

 The quantitative understanding of human dynamics is difficult and has not yet been explored in depth. The emergence of human interaction traces from online and pervasive environments allows us to understand details of human activities. For example, the Reality Mining project [7] collected proximity, location and activity information, with

Experimental data set	MIT	UCSD	CAM	INFC06	BATH
Device	Phone	PDA	iMote	iMote	PC
Network type	Bluetooth	WiFi	Bluetooth	Bluetooth	Bluetooth
Duration (days)	246	77	11	3	5.5
Granularity (seconds)	300	600	120	120	Continuous
Number of Experimental Devices	97	274	36	78	7431

Table 1: Characteristics of the experiments
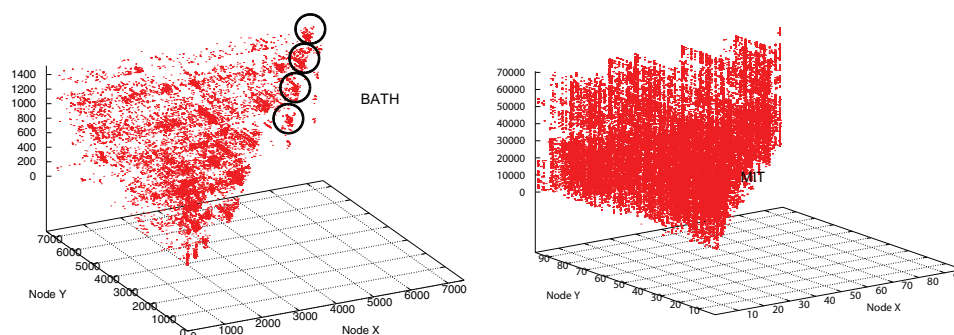
Fig. 1: Node Contact: BATH and MIT traces

nearby nodes being discovered through periodic Bluetooth scans and location information from cell tower IDs. Several other groups have performed similar studies. Most of these [7] [6] [22] use Bluetooth to measure device connectivity, while others [12] rely on WiFi. The duration of experiments varies from 2 days to over one year, and the numbers of participants vary. We have analysed various traces from the Crawdad database [3] listed below, and Table 1 summarises the configuration.

MIT: in the MIT Reality Mining project [7], 100 smart phones were deployed to students and staff at MIT over a period of 9 months. These phones were running software that logged contacts.

UCSD: in the UCSD Wireless Topology Discovery [29], approximately 300 wireless PDAs running Windows CE were used to collect WiFi access point information periodically for 11 weeks.

CAM: in the Cambridge Haggle project [18], 40 iMotes were deployed to 1st year and 2nd year undergraduate students for 11 days. iMotes detect proximity using Bluetooth.

INFC06: 78 iMotes were deployed at the Infocom 2006 conference for 4 days [2].

BATH: in the Cityware project, 9 Bluetooth scanners across the city of Bath were deployed to monitor the presence of mobile devices within an approximate 10 metre radius [23]. The co-location of a device pair is identified from the log data. Also part of devices are equipped with a Bluetooth scanning program [22] and detected device information is collected via GPRS. This leads to the construction of a connectivity graph for each time unit.

Note that it is a complex task to collect accurate connectivity traces using Bluetooth communication, as the device discovery protocol may limit detection of all the devices nearby. Bluetooth inquiry can only happen in 1.28 second intervals. An interval of $4 \times 1.28 = 5.12$ seconds gives a more than 90% chance of finding a device. However, there is no data available when there are many devices and many human bodies around. The power consumption of Bluetooth also limits the scanning interval, if devices have limited recharging capability. The iMote connectivity traces in Haggle use a scanning interval of approximately 2 minutes, while the Reality Mining project uses 5 minutes. The advantage of BATH data is that scanning is done continuously. The ratio of devices with Bluetooth enabled to the total number of devices is around 7%. Because of the uniqueness of urban-scale human connectivity data, we focus on analysis using the

BATH trace in this paper. Fig. 1 depicts all contact points between two nodes along the timeline in 3D form. The z-axis represents time, with 300 seconds per unit. This depicts the same node pair encountering repeatedly, which is marked with circles. The Bath data dictates 5 days repeating contact patterns, while the MIT trace shows as a vertical line during 9 months.

3   Complexity of Real World Networks

In general, to understand the network structure one requires three key metrics: the average path length to show the distance between a pair of nodes, the cluster coefficient to indicate how well nodes are clustered, and the degree distribution. In DTNs, the topology changes every time unit and data paths, which may not exist at any one point in time, potentially arise over time. Thus, existing metrics for static networks are difficult to apply. Previously, the characteristics of a pair of nodes – such as inter-contact and contact distribution – have been explored in several studies [2] to which we refer the reader for further background information. We also described the extraction of information related to levels of clustering or network transitivity, and strong community structure in our previous work [34] [14].

As PSNs are formed by humans, it is assumed that social networks take a major role in epidemic spread. Most social networks are neither random nor regular but complex. The properties of nodes include fixed states, variable states, neighbour nodes and network positions (i.e. centralities). Understanding a complex system requires not only understanding of the elements in the system, but also of the patterns of interactions between the elements. Thus, observing communication over the network is expected to give some information about the network structure and, vice versa, the network structure affects the communication. In this paper, we focus on information flow during epidemic spread, including the impact of hub nodes. In the following subsections, we discuss various metrics that can be used in expressing dynamic time-dependent networks.

3.1   Node Distance and Clustering

The average shortest path length between any two randomly chosen people on the planet (i.e. 6.5 billion people) is 6. This is easy to explain if social ties are highly random. However, real social networks are not random, as they exhibit a great deal of clustering, and the average distance between two nodes is small. There are also shorcuts between clustered groups. A network with small average degrees, high clustering, and small average distances has been called a small world network by Watts [31]. Table 2 summarises the average hop counts and cluster coefficient values for each trace. The cluster coefficient value of the MIT trace – 0.44 – is the probability that, if node A knows nodes B and C, nodes B and C know each other. The BATH trace, where proximity data is collected in city scale, shows an average hop count of 3.3 and cluster coefficient value of 0.45.

Experimental traces	Average Hop Count	Cluster Coefficient
MIT	1.6	0.44
UCSD	2.2	0.41
CAM	1.2	0.66
INFC06	1.5	0.52
BATH	3.3	0.45

Table 2: Average Hops and Cluster Coefficient

Time unit = t

Time unit = t+1

Time unit = t+2

● Node
— High weight edge
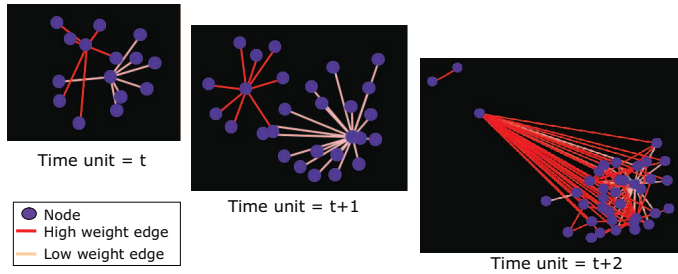— Low weight edge

Fig. 2: Evolution of Connection Map and Edge Characteristics (UCSD Trace)

3.2 Weighted Graph

The connectivity traces can be represented by weighted graphs – also called contact graphs – in which the weight of an edge represents the *contact duration* and *contact frequency* for the two end vertices. Understanding human interaction can then be tackled in the domain of weighted network analysis. Possible outcomes from studying of the weighted contact graphs include community detection and determining node centrality. Many real world networks are weighted, but due to complexity, little analysis has been done in this area. The seminal work is a weighted network analysis paper by Newman [20]. A weighted graph can be converted into a multi-graph with many unit edges. Here, we only consider symmetric edges. In reality, edges can be symmetric (undirected) or asymmetric (directed), possibly with a different strength in either direction. Fig. 2 depicts network evolution over a period 15 minutes in the UCSD trace (taken from our visualization work [35]). The network exhibits a small-world-like formation at first, which breaks down into two groups, each forming a star topology. See Section 6 for further community detection.

3.3 Node Centrality

Understanding a network and a node's participation in the network is important. Centrality measurements give insight into the roles and tasks of nodes in a network. The centrality of a node in a network is a measure of the structural importance of that node. Freeman defined several centrality metrics [10] and three of the best-known metrics are described below:

Degree centrality C_D of a node a measures the number of direct connections d. It indicates how active a node is in the network.

$$C_D(a) = \ d_a \tag{1}$$

Social networks in general exhibit small average degree compared to the number of nodes, where people have limited connections to the other people. There are over 200 million web sites, with an average degree of only 7.5, and most sites with less than 10 links, but some sites have thousands of links. In time-dependent networks, the degree centrality should ideally be calculated within an appropriate time-window (see further discussion in Section 5). Fig. 3 depicts the degree distribution of BATH, INFC06 and UCSD traces: the BATH and INFC06 traces exhibit a power-law distribution, whereas the UCSD trace shows that most nodes have a similar degree. The UCSD data is based WiFi and may not have as precise proximity information unlike the other traces. Fig. 4
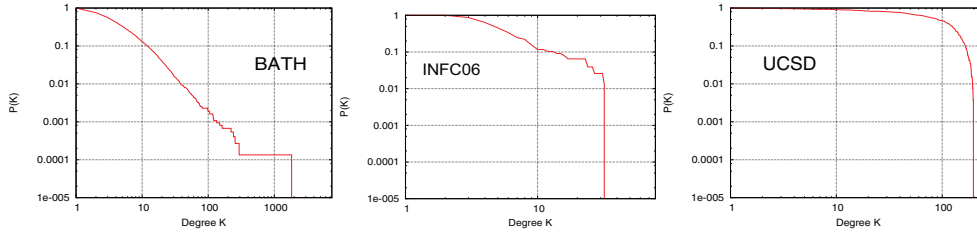
Fig. 3: Aggregated Degree Distribution
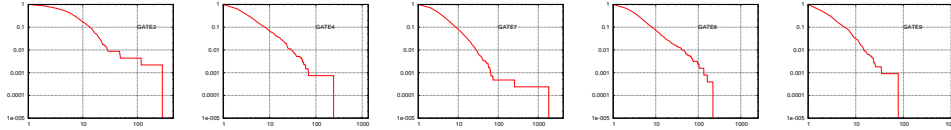
Fig. 4: Degree Distribution: Gates in Bath Trace

shows the degree distribution at the scanner locations in the BATH trace that exhibit power-law distribution.

Betweenness centrality C_B indicate that a node acts as a bridge between two nonadjacent nodes. Thus, a node with high betweenness potentially has control over these two nonadjacent nodes. A high-betweenness node in the network may impact on the data flow between two groups of nodes.

$$C_B(a) = \sum_{b<c}[g_{bc(a)}/g_{bc}] \tag{2}$$

where g_{bc} is the number of geodesics between b and c, and $g_{bc(a)}$ is the number of geodesics between b and c that contain a. In other words, the betweenness centrality is a sum over all pairs (b, c) of the proportion of geodesics linking the pair that contain node a. Betweenness centrality in time-dependent networks may be calculated using traffic simulation to establish the role of each node (see further discussion in Section 5).

Closeness centrality C_C indicates the visibility of a node in the network and sub-network. Maximising closeness centrality yields the node with the shortest path to all others and the best visibility. We have used closeness centrality to build an overlay over the communities [34]. It is a measurement of how long it will take data to spread the others in the community. The closeness centrality, $C_C(a)$, for a vertex a is inverse of the sum of distances to all other nodes:

$$C_C(a) = 1 / \sum_b d_{ab} \tag{3}$$

3.4 Dynamic Human Behaviour

Analysing the structural properties of growing networks could be relevant for social networks. In each time unit t_i, several nodes appear or disappear, and each selects or deselects k possible counter parts from the existing networks. They join or leave the
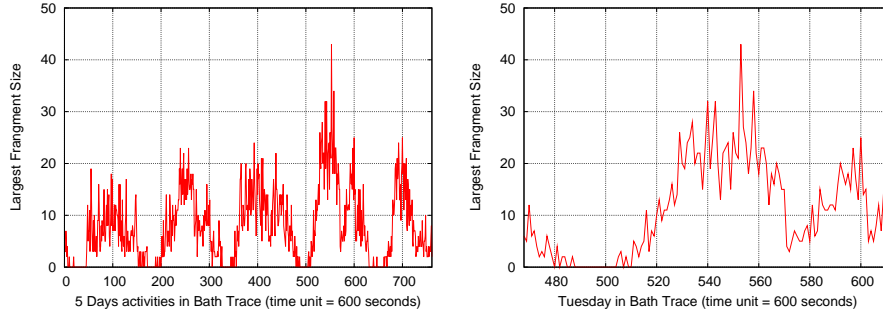
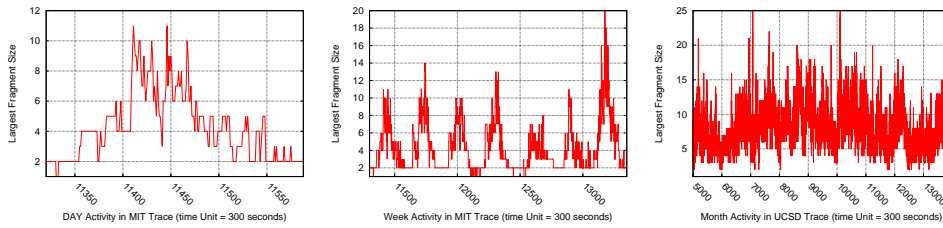Fig. 5: Largest Fragment in Timeunit (Bath Trace)

Fig. 6: Largest Fragment in Timeunit (MIT and UCSD Traces)

network with probability p. Identifying the values k and p from the empirical trace defines the form of network evolution. When p is large, over many time steps the network transition is significant.

Fig. 5 depicts the size of the largest connected subgraph in each time unit, based on the BATH trace, which shows the network dynamics over **4 days (Sunday through Wednesday)**. The snapshot on Tuesday depicts a single day's activity, and distinct day- and night-time dynamics can be observed. Fig. 6 shows the same dynamics, including monthly periodicity, based on the MIT and UCSD traces. Note that the size of the largest fragment in the BATH trace is slightly larger than in the MIT/UCSD traces, because the BATH trace covers all devices in the city of Bath, wheres the MIT/UCSD traces only consider a known group of 100–270 participants. The cause of larger fragment sizes in the BATH trace raises an interesting question: are these due to temporal/spatial connections or tighter social connections?

4   Epidemic Dynamics

Epidemiology can be used to deal with intermittent connectivity in DTN environments. The small-world topology of interpersonal connection and its hierarchical structure yields a two-level structure that has a strong impact on epidemic spread in a population. DTNs bring a further complex new network structure, because devices can either communicate through the communication mechanism like the Internet, or directly when they are in the communication range using short-range wireless communication.

Pastor-Satorras has conducted an analytical and numerical study on a large-scale dynamical model on epidemic spread in synthetic networks [26][25]. In this section, we show various epidemic characteristics from our experiments using the real world traces.
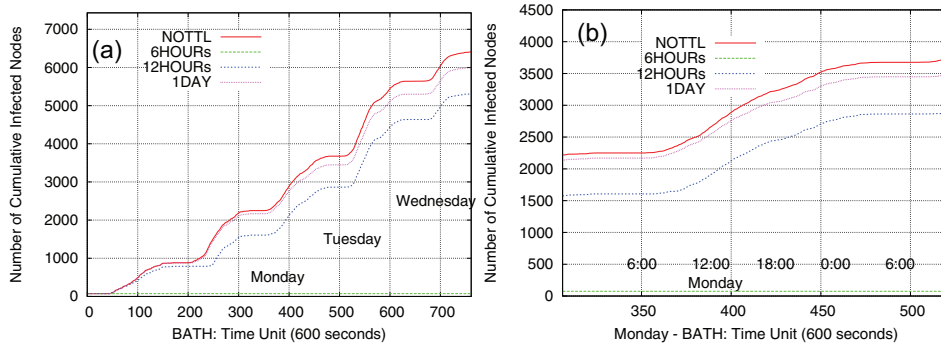
Fig. 7: BATH: Epidemic Spread

4.1   SIS Model

For epidemic spread, we use the *Susceptible-Infected-Susceptible (SIS)* model. Each node in the network represents an individual, and each link is a connection along which the virus infection can spread from between individuals. The SIS model is defined as follows:

1. Each node can be in one of two states:
   – Susceptible (not currently infected)
   – Infectious (infected)
2. The initial infectious nodes may be drawn from the following groups. These nodes do not participate in epidemic propagation until they appear in the trace.
   – Top percentile of the degree distribution
   – 50th percentile of the degree distribution (i.e. average)
   – Bottom percentile of the degree distribution
3. When a node is infectious, it can infect the other nodes with probability λ, where $\lambda = 1$. At each time unit, if that node has a link with a susceptible node, the susceptible node becomes infectious.
4. In each infectious node, the virus has a time-to-live (TTL). When the TTL expires, the node reverts to the susceptible state.

Fig. 7 depicts how the infectious nodes change with time, based on the BATH trace, and using the nodes in the top percentile of the degree distribution as the original infectious nodes. When the TTL is set to 6 hours, we do not observe epidemic spread
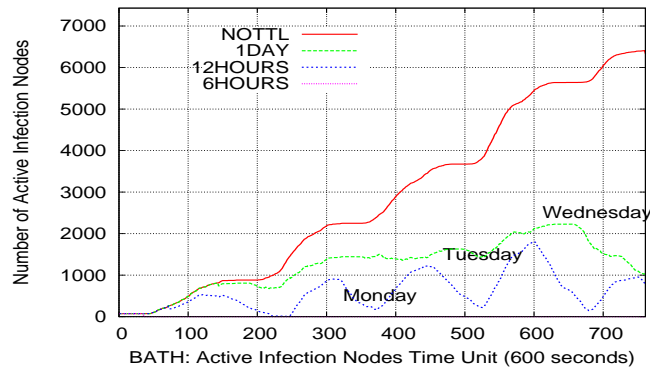

Fig. 8: BATH: Active Infected Nodes

at all; while, if the TTL is set to 1 day, the effect is similar to having an infinite TTL. Fig. 8 depicts the number of infectious nodes during epidemic spread. With a TTL of 12 hours, a circadian cycle can be observed, with an increase in the number of infectious nodes during day-time and the virus dying out at night. However, the resilient epidemic comes back during the next day. The trace is not long enough to see the trend towards the end of Wednesday and we plan to conduct extended experiments that will yield traces ranges from a month to a year in length.

We have conducted experiments to investigate the impact of selecting different source nodes. As stated above, we base our selection on the distribution of node degrees. When the bottom percentile of this distribution (75 nodes) is selected to give source nodes, epidemic spread only begins after 1 day, whereas starting with high degree nodes causes epidemic spread to begin immediately. Once the epidemic spread has begun, the spread proceeds at a similar rate in either case.

Fig. 7(b) shows three stages of epidemic spread during a 24-hour period. The stages are (1) a rapid increase at first where propagation may take place within clusters, (2) slow climbing when infectious nodes encounter external clusters, and (3) exhaustion of infection as the epidemic spread hits the upper limit of infection. During the first stage, linking between clusters may occur, and this accelerates the increase of infected nodes. This three stages can be observed in the MIT and UCSD traces (see Fig. 9 and Fig. 10). Fig. 9(b) and Fig. 10(b) depict the second stage with an enlarged time unit scale.
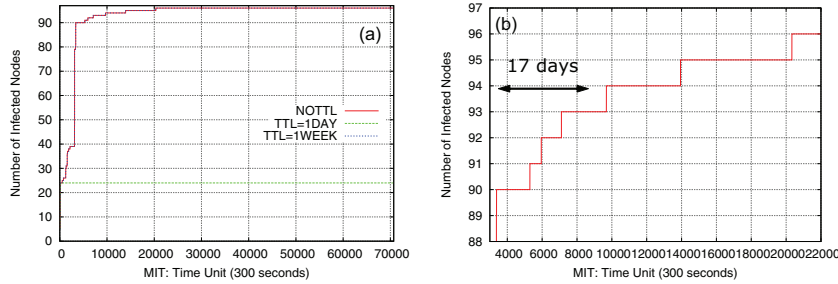

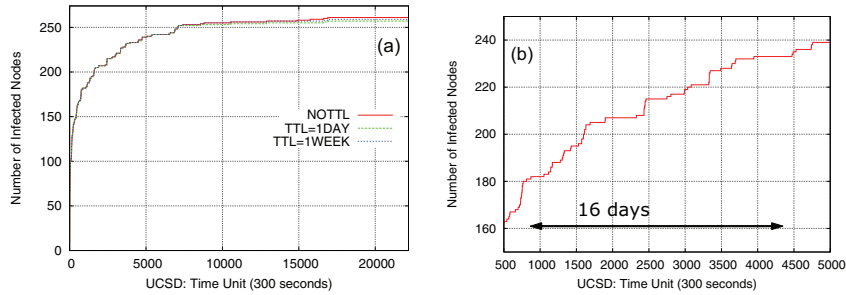**Fig. 9: MIT: Epidemic Spread**

Fig. 10: UCSD: Epidemic Spread

5 Influence of Hub Nodes

In this section, we investigate hub nodes and their influence on epidemic spread using the BATH trace. We have defined hubs based on the following centralities and extracted
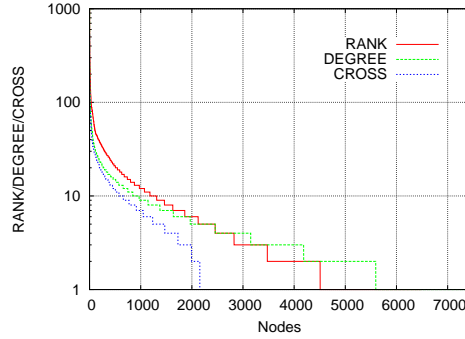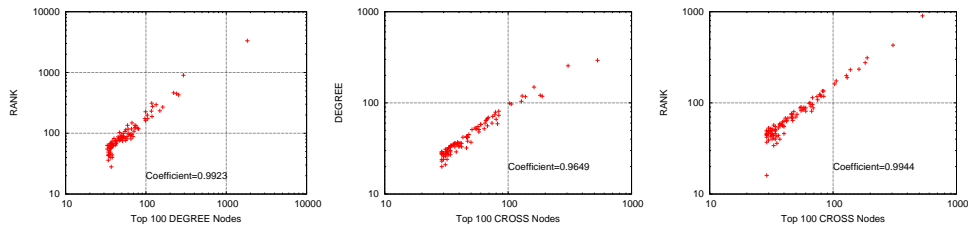
Fig. 11: Hub Nodes

Fig. 12: Correlation of 100 Hub Nodes

the top 100 such hub nodes from the trace. We then ran the epidemic spread simulation described in Section 4 but excluding the hub nodes, in order to observe how much influence they have on the spread.

DEGREE Hub: The total degree of each node over the entire duration of the trace indicates the popularity of the node (*Degree Centrality*). With this metric, it is not possible to distinguish two types of hubs: the node has high degree within a short time window (*party hub*) or a larger time window (*date hub*) [11]. Most nodes interact with only a few other nodes while a small number of hub nodes may have many interactions.

In [13], we examined the degree per unit time (e.g. the number of unique nodes seen per 6 hours). We chose a 6-hour time window based on our intuition that daily life is divided into 4 main periods: morning, afternoon, evening and night. This is similar to the approach described in [33]. However, it is sensitive to starting the time window at different absolute times of the day. As Fig. 5 shows, the day cycle could be a more efficient time window in the urban space.

RANK Hub: The frequency that a node is used to relay data to other nodes indicates the centrality of the node. We simulated flooding over the temporal graph extracted from the trace and counted the number of times each node is used for relaying the data. We exploited different counting schemes, such as counting any time a node relays data or only when the node is on the shortest path from the source to the destination. Different schemes result in a similar ranking. This metric is equivalent to *Betweenness Centrality* in time-dependent networks.

CROSS Hub: The appearance of a node at different locations indicates that it has *Mobility Centrality*. With the BATH trace, 9 locations are extracted and the rate of ap-

pearance at each location is measured. Fig. 11 depicts the distribution of all nodes with extracted centrality metrics. The y-axis shows a centrality metric on a logarithmic scale.

5.1 Hub Nodes Similarity

Fig. 12 and Table 3 show the correlation between the sets of hub nodes using different metrics. The coefficient values are greater than 0.95 in every case. The correlation between *RANK HUB* and *CROSS HUB* has the highest value. Table 4 depicts the membership similarity of hub nodes. The *RANK HUB* and *DEGREE HUB* sets share many nodes, while the *CROSS HUB* set has only around 50% of nodes in common with the *DEGREE HUB* or *RANK HUB* sets.

Category	All Nodes	Top 100 Nodes	Top 50 Nodes	Top 30 Nodes
Rank/Degree	0.99	0.99	0.99	0.99
Degree/Cross	0.97	0.96	0.96	0.96
Corss/Rank	0.99	0.99	0.99	0.99

Table 3: Hub Nodes Correlation

Top n Nodes	Rank/Degree	Rank/Cross	Degree/Cross
100	0.79	0.43	0.44
70	0.92	0.41	0.41
50	1.00	0.43	0.49
30	1.00	0.46	0.46
10	1.00	0.33	0.33

Table 4: Hub Node Membership Similarity

5.2 Inactivation of Hub Nodes

Fig. 13 depicts the impact of deactivating hub nodes during the epidemic spread. Fig. 13 shows inactivation of the top 50 nodes. Removing the top 50 *DEGREE HUB* or *RANK HUB* nodes significant reduces the epidemic spread. Both *DEGREE HUB* and *RANK HUB* nodes have a similar impact. On the other hand, removing the *CROSS HUB* nodes does not show as dramatic an impact as does removing the other two types of hub nodes. Randomly selected 1% of top 30% of high degree nodes are used as the source
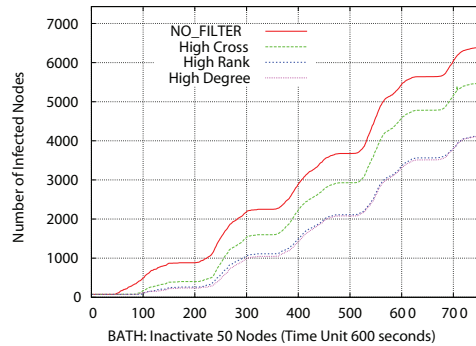
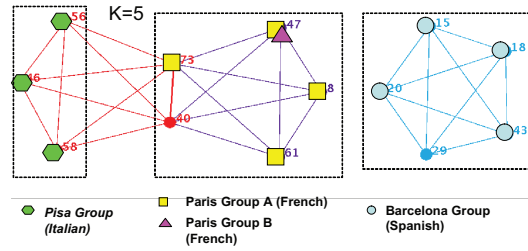**Fig. 13: Inactivation of Hub Nodes**

Fig. 14: k-CLIQUE Community Detection in INFC06 Trace

of the infection. The result indicates the strong influence of hub nodes. We are further investigating what differentiates static hubs from dynamic hubs in a pair interaction.

6   Inferring Human Communities

People inherently form groups, yielding social structures in which prominent patterns or information flow can be observed. We have worked on uncovering the structure and dynamics of social communities from human connectivity traces, in which social groups must be embedded [14] [34]. We have shown various community detection mechanisms which can be applied to human connectivity traces in both a centralised and a decentralised way.

Community detection in complex networks has attracted a lot of attention in recent years. In the Internet, community structures correspond to autonomous systems. It is crucial to construct efficient algorithms for identifying the community structure in a generic network. Many community detection methods have been proposed and examined in the literature (see the recent review papers by Newman [21] and Danon *et al.* [5]).

We have exploited different algorithms [13]. The k-CLIQUE method has been designed for binary graphs, and we therefore need to threshold the edges of the contact graphs in the traces [24], while *Weighted Networks Analysis* [20] can work on weighted graphs directly without any threshold.
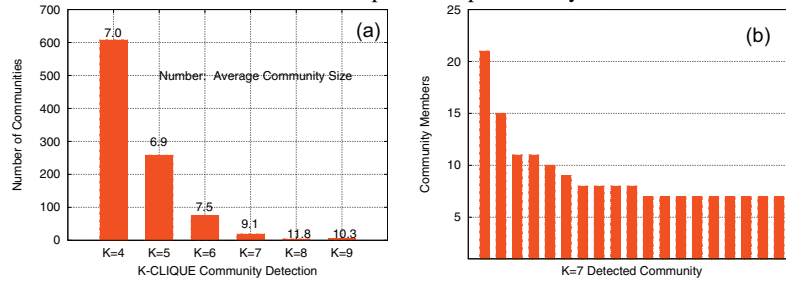
6.1 K-CLIQUE Community Detection

Palla *et al.* define a community as a union of all k-cliques (complete sub-graphs of size k) that can be reached from each other through a series of adjacent k-cliques, where two k-cliques are said to be adjacent if they share $k-1$ nodes. As k is increased, the k-clique communities shrink, but on the other hand become more cohesive since their member nodes have to be part of at least one k-clique. An advantage of this approach is that it allows overlapping communities, which is useful as, in human society, one person may belong to multiple communities.

Fig. 14 depicts the detected communities in the INFC06 trace, when $k = 5$. Three distinct communities are detected, which include two nodes that belong to two communities. Fig. 15(a) shows the detected communities with different k values and Fig. 15(b) depicts the community size distribution when $k = 5$ in the BATH trace.

6.2   Inter- and Intra-Gate Communities

It is known that the location is an important attribute for social community structure. The BATH trace includes the location of scanners (i.e. Gate 1 to 9), and Fig. 16 depicts

**Fig. 15: Community Detection in Bath City Trace**

the appearance rate of community members at Gates 2, 4, 7 and 8. For example, the members of community 17 (with solid circle) are observed at Gate 2, but almost never at the other locations; whereas the members of communities 1 and 15 (with dashed circle) appear at every gate. We refer to the former type of community as an *Intra-Community*, and to the latter type as an *Inter-Community*. Intra-communities may have a strong tie with the location, while inter-communities may indicate a group of people moving together.

We ran a simulation to investigate the effect of deactivating 100 nodes of each community type. Communities 1 and 15 were selected at random to represent inter-communities, and 7 and 17 are selected as intra-communities. Removing intra-communities causes an up-to-10% reduction in infectious nodes, while removing inter-communities has no effect in this scenario. It is well known in social networks that inter-relationship within a group is stronger than external links. The experiment result indicates the characteristics of social networks.

The communities detected in the traces may be static social communities or transient communities, such as a group of people who happen to be in the same location. Our current approach does not distinguish between these two different community concepts and further refinement of community concepts, along with membership management is part of our ongoing work.

7   Related Work

The recent discovery of complex network properties in the structure of biological and social systems [28] has brought different perspectives on real world networks. Traditionally, random networks have been studied extensively [8]. Random graphs are usually constructed by randomly adding links to a static set of nodes. Random graphs tend to have short paths between a pair of nodes. Recent work on random graphs has provided mechanisms to construct graphs with specified degree distributions. Power-law networks are networks where the probability that a node has a degree k is proportional to k. Many real-world networks have been shown to be power-law networks, including Internet topologies and social networks. Scale-free networks are a class of power-law
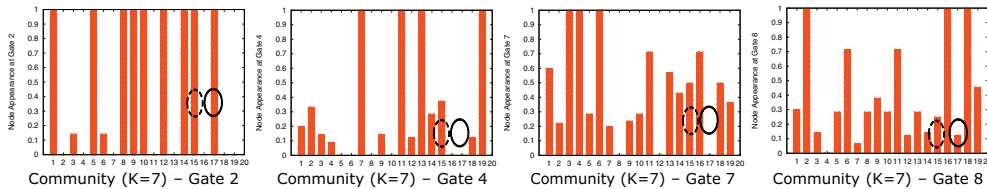
Fig. 16: Bath Trace: Communities in Gates

networks where the high-degree nodes tend to be connected to other high degree nodes. Small world networks have a small diameter and exhibit high clustering [31]. Studies have shown that the Web, scientific collaboration on research papers, film actors, and general social networks have small world properties []. It has become clear that this pattern of interactions, which forms the network, plays a fundamental role in understanding these systems.

Most forwarding algorithms in DTNs are based on epidemic routing protocols [30], whereby messages are simply flooded when a node encounters another node. The optimisation of epidemic routing by reducing the number of copies of a message has been explored. Many approaches calculate the probability of delivery to the destination node, where the metrics are derived from the history of node contacts, spatial information and so forth. The pattern-based Mobyspace Routing by Leguay *et al.* [17], location-based routing by Lebrun *et al.* [16] and PROPHET Routing [19] fall into this category. The Message Ferry approach of Zhao *et al.* [32] takes a different approach by controlling the movement of each node. Recently, attempts to uncover hidden stable network structure in DTNs and social networks have emerged. For example, SimBet Routing [4] uses ego-centric centrality and its social similarity. Messages are forwarded towards a node with higher centrality to increase the possibility of finding the potential carrier to the final destination.

Emerging wireless technologies are creating physical network in the actual physical space along online communication (e.g. social network services, email). Understanding this new pervasive network as a time-dependent dynamic human network is still an open research area. Social relationships and interactions (i.e. social context) is gaining importance. New results in the area of complex network theory [1] give new insight on social networks.

8   Conclusions and Future Work

In this paper, we have shown our study of epidemic spread in dynamic human networks from human connectivity traces. The human networks exhibit periodic activity. Daily circulation is significant, and epidemic spread demonstrates that if the virus has over one day of life, the spread rate reaches almost the same level as when the virus has infinite life. Removing the top 100 hub nodes (using various definitions of "hub") out of over 7500 nodes from consideration yields a significant reduction in the rate of epidemic spread.

In the BATH trace, some communities exhibited strong ties to particular locations. Therefore the local network structure could possibly form a type of small-world network, and a small number of nodes could connect with external nodes, forming a scale free network. We are currently working to prove this assumption by constructing corresponding synthetic networks that can be compared with real world networks. Our future work includes investigating an asymmetric communication model (i.e. forming a directed graph) and defining new network measurement criteria such as time-dependent centralities and cluster coefficient values. We are taking an empirical approach and therefore obtaining accurate and fine grained trace data is essential. We are planning to deploy several urban scale experiments for data collection and information diffusion.

Pervasive DTNs are dynamic, and we are particularly interested in how network structure affects information flow, and vice versa: how the ongoing communication af-

fects the network structure. Pairwise communication and social structure need to be integrated and modelled alongside dynamic interactions. The social network reflects access to information and change of social activities can be seen as seeking better information access. Our ultimate goal is a complete understanding of human-to-human network models in the urban space.

References

1. R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47, 2002.
2. A. Chaintreau et al. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proc. INFOCOM*, April 2006.
3. D. College. A community resource for archiving wireless data at dartmouth, http://crawdad.cs.dartmouth.edu/index.php, 2007.
4. E. Daly and M. Haahr. Social network analysis for routing in disconnected delay-tolerant manets. In *Proceedings of ACM MobiHoc*, 2007.
5. L. Danon, J. Duch, A. Diaz-Guilera, and A. Arenas. Comparing community structure identification, 2005.
6. C. Diot et al. Haggle Project, http://www.haggleproject.org, 2008.
7. N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, V10(4):255–268, May 2006.
8. P. Erdos and A. Renyi. On random graphs i. *Mathematicae*, 5, 1959.
9. K. Fall. A delay-tolerant network architecture for challenged internets. In *Proc. SIGCOMM*, 2003.
10. L. C. Freeman. A set of measuring centrality based on betweenness. *Sociometry*, 40:35–41, 1977.
11. J.-D. J. Han1, N. Bertin1, T. Hao1, D. S. Goldberg, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430, 2004.
12. T. Henderson et al. The changing usage of a mature campus-wide wireless network. In *Proc. Mobicom*, 2004.
13. P. Hui, J. Crowcroft, and E. Yoneki. BUBBLE Rap: Social Based Forwarding in Delay Tolerant Networks. In *MobiHoc*, 2008.
14. P. Hui, E.Yoneki, S. Chan, and J. Crowcroft. Distributed community detection in delay tolerant networks. In *Proc. MobiArch*, 2007.
15. J. Kleinberg. The wireless epidemic. *Nature*, 449(20), 2007.
16. J. Lebrun, C.-N. Chuah, et al. Knowledge-based opportunistic forwarding in vehicular wireless ad hoc networks. *VTC 2005*, pages 2289–2293, 2005.
17. J. Leguay et al. Evaluating mobility pattern space routing for DTNs. In *Proc. INFOCOM*, 2006.
18. J. Leguay et al. Opportunistic content distribution in an urban setting. In *ACM CHANTS*, 2006.
19. A. Lindgren, A. Doria, and O. Schelen. Probabilistic routing in intermittently connected networks. In *Proc. SAPIR*, 2004.

20. M. Newman. Analysis of weighted networks. *Physical Review E*, 70:056131, 2004.
21. M. Newman. Detecting community structure in networks. *Eur. Phys. J. B*, 38:321–330, 2004.
22. T. Nicolai, E. Yoneki, N. Behrens, and H. Kenn. Exploring social context with the wireless rope. In *Proc. Workshop MONET: LNCS 4277*, 2006.
23. E. O'Neill et al. Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In *UbiComp*, pages 315–332, 2006.
24. G. Palla et al. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.
25. R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E.*, 64(066117), 2001.
26. R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scalefree networks. *Phys. Rev. Lett.*, 86(14), 2001.
27. S. J. Rahul C Shah, Sumit Roy and W. Brunette. Data mules: Modeling a three-tier architecture for sparse sensor network. In *IEEE Workshop on Sensor Network Protocols and Applications (SNPA)*, May 2003.
28. S. H. Strogatz. Exploring complex networks. *Nature*, 410 (268-276), 2001.
29. UCSD. Wireless topology discovery project, http://sysnet.ucsd.edu/wtd/wtd.html, 2004.
30. A. Vahdat and D. Becker. Epidemic routing for partially connected ad hoc networks. Technical Report CS-200006, Duke University, April 2000.
31. D. J. Watts. *Small Worlds – The Dynamics of Networks between Order and Randomneess*. Princeton University Press, Princeton, New Jersey, 1999.
32. M. A. Wenrui Zhao and E. Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *ACM Mobihoc*, May 2004.
33. P. Winters. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6:324–342, 1960.
34. E. Yoneki, P. Hui, S. Chan, and J. Crowcroft. A socio-aware overlay for multi-point asynchronous communication in delay tolerant networks. In *Proc. MSWiM*, 2007.
35. E. Yoneki, P. Hui, and J. Crowcroft. Visualizing Community Detection in Opportunistic Networks. In *ACM MobiCom - CHANTS*, 2007.