Distinct Types of Hubs
in Human Dynamic Networks

Eiko Yoneki, Pan Hui and Jon Crowcroft
University of Cambridge, Computer Laboratory
[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 using dynamic interconnectedness as people encounter each other lead towards a world where digital traffic flows. We focus on human-to-human communication in such environments, which exhibits the characteristics of social networks. Among various social contexts, we examine hubs in dynamic human networks. Hubs are highly connected nodes within the network. This paper describe hubs in time-dependent networks and two distinct types of hubs: *party* and *date* hubs from real world connectivity traces.

1. INTRODUCTION

Increasing numbers of mobile computing devices form dynamic networks in everyday life. In such environments, nodes (i.e. laptops, PDAs, smart phones) are sparsely distributed forming a network which is often partitioned due to geographical separation or node movement. We envision new communication paradigms using dynamic interconnectedness between people and urban infrastructure that lead towards a world where digital traffic flows as people pass each other [13]. Delay Tolerant Networks (DTNs) [7] aims 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: 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. To reduce the overhead of epidemic routing, various approaches have been reported ranging from count, timer, history based controlled flooding to location-based strategies.

We have shown an approach using a logical connection topology, where we uncovered a hidden stable network structure such as social networks [12] [21] from the human connectivity traces, which consist of a group of people forming socially meaningful relationships. In PSNs, social networks could map to computer networks since people carry the computer devices. We have shown improved performance with controlled epidemic strategy by applying extracted social contexts [11]. During the above work, we realised that further understanding network structure is essential since 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. Understanding of 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, and the interaction patterns are complex. This paper studies the network structure in such complex dynamic human networks focusing on the characteristics of hub nodes. Many studies for network modelling have been reported, which are based on simulation or small collection of data. We emphasise here the use of real world data and believe that our study will also provide interesting insight into the real human interaction. We consider a model for time paths based on graph evolution: *Time-Dependent Networks* where links between nodes are time-windows dependent. We report preliminary empirical results in this paper.

The rest of this paper is structured as follows. We introduce the experimental data sets in Section 2, followed by describing complexity of real world connectivity data in Section 3. We discuss hub nodes in Section 4 and party and date hubs in Section 5 followed by related work. Finally we conclude the paper with a brief discussion.

2. HUMAN CONNECTIVITY TRACES

The quantitative understanding of human dynamics is difficult and has not been explored in depth. The emergence of capturing traces of human interaction in online and pervasive environments allows us to understand details of human activities. For example, the Reality Mining project [5] collected proximity, location and activity information, with 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 [5] [4] [16] use Bluetooth to measure device connectivity, while others [10] rely on WiFi. The duration of experiments varies from 2 days to over 1 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 [5], 100 smart phones

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 Devices	97	274	36	78	7431

Table 1: Characteristics of the experiments

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

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 [19], approximately 300 wireless PDAs running Windows Pocket PC were used collecting WiFi access points information periodically for 11 weeks.

CAM in the Cambridge Haggle project [14], 40 iMotes were deployed to 1st and 2nd year undergraduate students for 11 days. iMote detects proximity by Bluetooth.

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

BATH: 9 Bluetooth scanners across the city of Bath are deployed to monitor the presence of mobile devices within an approximate 10 metre radius in the Cityware project [17]. Collocation of a pair devices is identified from the log data leading to construction of connectivity graph in each timeunit. The scanners collect 7431 nodes for 5.5 days.

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. 4×1.28 (i.e. 5.12 seconds) gives you more than 90% chance of finding a device. However, there is no data available for many devices and many human bodies. Power consumption of Bluetooth also limits scanning interval if devices have limited recharging capability. iMote connectivity traces in Haggle use around 2 minutes scanning interval, while the Reality Mining project uses 5 minutes. The advantage of BATH data is scanning is done continuously. The propotion of Bluetooth being turned on for the device in the city of Bath is around 7%.

3. COMPLEX REAL WORLD NETWORKS

In general, to understand the network structure requires three key metrics: the average path length to show the distance of a pair of nodes, the cluster coefficient to indicate how well nodes are clustered, and the degree distribution. Table 2 summarises the average hop counts and cluster coefficient values. In DTNs the topology changes every timeunit where data paths may not exist at any one point in time but potentially do exist over time. Thus, existing metrics for static networks are difficult to apply. Previously the characteristics of a pair of nodes have been explored in several studies [2] such as inter-contact and contact distribution to which we refer for further background information. We also described extracting information related to levels of clustering or network transitivity and strong community structure in our previous work [21] [11] [12].

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). A complex system requires not only understanding, the elements in the system but the interactions and patterns between the elements is also important. Thus, observing communication over the network is expected to give some guide to infer the network structure and, vice versa, the network structure

affects the communication.

3.1 Weighted Graph

The connectivity traces can be represented in the form of weighted graphs called contact graphs, with the weight of an edge representing the *contact duration* and *contact frequency* for the two end vertices. Hence understanding human interaction can be tackled from the domain of weighted network analysis. Possible outcomes from studying of the weighted contact graphs include community detection and weighted node centrality. Many real world networks are weighted, but because of their complexity, little analysis has been done in this area. The seminal work is a weighted network analysis paper by Newman [15]. A weighted graph can be converted into a multi-graph with many unit edges. In this paper, we only consider symmetric edges. In reality, edges can be symmetric or asymmetric (un-directional or directional) and can have different strength. Fig. 1 depicts network evolution for 15 minutes in the UCSD trace from our visualization work [22]. The network forms a small world like formation at the beginning, which was broken into two groups of nodes forming star topology.

3.2 Dynamic Human Behaviour

Analysing the structural properties of growing networks could be relevant to social networks. At each timeunit n, a number of new nodes appear/disappear and select/deselect k possible counter parts from the existing networks. They join/disjoin the network with probability p. Identifying the values k and p from the empirical trace defines the network evolution form. When n and p are large, network transition is significant. Fig. 2 depicts the largest fragment size in each timeunit extracted from the BATH trace, which shows the dynamics over 5 days of activities from Sunday through Wednesday. The snapshot on Tuesday depicts a day activity, where day and night dynamics are observed.

4. HUB NODES

Understanding nodes 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 the node. Freeman [8] defined several centrality metrics and three well known centrality measures are *Degree, Betweenness,* and *Closeness*. *Degree* and *Betweenness* centralities are described below:

Degree Centrality: $C_D(a) = d_a$, where C_D on a node a measures the number of direct connections d. This indicates that the
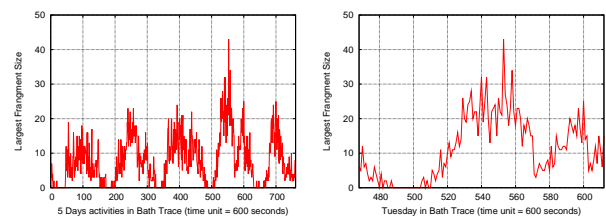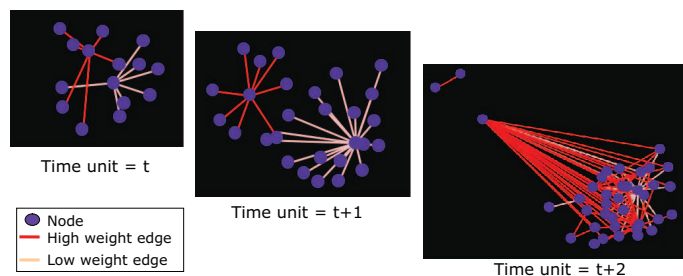

Figure 2: Largest Fragment in Timeunit (Bath Trace)

Figure 1: Evolution of Connection Map and Edge Characteristics (UCSD Trace)

node must be the most active in the network.

Social networks in general exhibit a small average degree compared to a number of nodes, and there is huge inequality. On the web, in a study from 2006, there were over 200 million web sites, average degree only 7.5, most web sites less than 10 links, but some has thousands of links. In time-dependent networks, ideally *Degree* centrality should be calculated within an appropriate time-window. Fig. 3 depicts the degree distribution of BATH, INFC06 and UCSD traces, where the BATH and INFC06 traces exhibit a power-law distribution, while the UCSD trace shows that most nodes have similar degrees. The UCSD data is WiFi based, which may not have precise proximity information as the other traces have.

Betweenness Centrality: $C_B(a) = \sum_{b<c}[g_{bc(a)}/g_{bc}]$, where C_B indicates a bridge node between two nonadjacent nodes. 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, it is a sum over all pairs (b, c) of the proportion of geodesics linking the pair that contains node a. Thus, a high betweenness potentially might have control over these two nonadjacent nodes. A betweenness node in the network may impact data flow between two groups of nodes.

We have defined the following hubs for the *time-dependent network* based on the above concept of centralities.

DEGREE Hub: is the total degree of each node in entire duration of the trace that 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 or a larger time window. Most nodes interact with only a few other nodes, while a small number of hub nodes may have many interactions. See Section 5 for further distinction of such types of hubs. In [11], we have examined the degree per unit time such as the number of unique nodes seen per 6 hours. We experimented with 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 a similar approach as described in [20]. However, it is sensitive to select the time window at the absolute time of the day. As Fig.2 shows, the day cycle could be a more efficient time window in the urban space.

RANK Hub: indicates how often each node is used to relay data to other nodes. We simulate flooding over the temporal graph extracted from the trace and counted the number of times each node
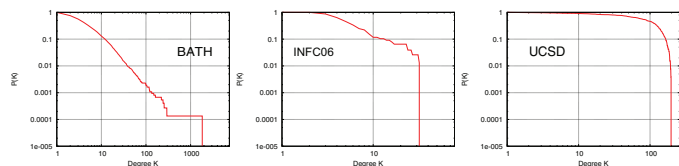
used for relaying the data. We have exploited different counting algorithms such as counting any time the node replays data or counting when the node is used as part of the shorted path from the source node to the destination node. Different algorithms result in a similar ranking. This metric indicates *Betweenness Centrality* in time-dependent networks.

CROSS Hub: defines how often a node appears at different locations that indicates *Mobility Centrality*. With the BATH trace, 9 locations are extracted, and the rate of appearance at an individual location is measured. Fig. 5 depicts the distribution of all the nodes with extracted centrality metrics. Y axe shows a log scale of centrality metrics.

We extracted the top 100 centrality nodes from the trace and explored the influence by such hub nodes.

4.1 Hub Nodes Similarity

Fig. 4 and Table 3 show the correlation of hub nodes. The coefficient values are high over 0.95 in every case. The correlation between *RANK HUB* and *CROSS HUB* shows the highest value. Table 4 depicts the membership similarity of hub nodes. *RANK* and *DEGREE* hubs share many nodes, while *CROSS* share only around 50% of nodes with either *DEGREE* or *RANK* hub nodes.

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
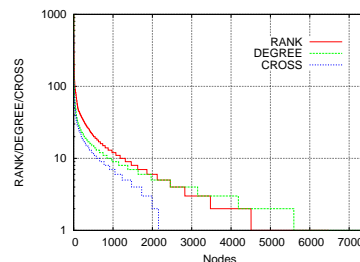

Figure 3: Aggregated Degree Distribution
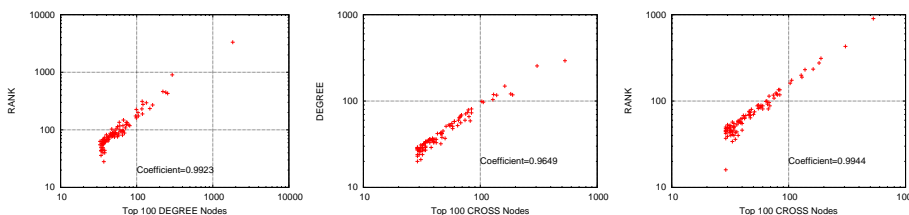

Figure 5: Hub Nodes

9

Figure 4: Correlation of 100 Hub Nodes

4.2 Impact of Hubs in Epidemic Spread

We have exploited experiments to see the impact of hub nodes during the epidemic spread using the human connectivity trace (see [23] for the details). The top 100 of each type of hub node are inactivated during the epidemic spread. The disappearance of *DEGREE HUB* or *RANK HUB* completely killed epidemic spread. Both *DEGREE* and *RANK* hubs show a similar impact. On the other hand, *CROSS HUB* does not show as dramatic as impact as the other two types of hub nodes do. Fig. 6 depicts an impact of inactivation of top 100 hub nodes during the epidemic spread.
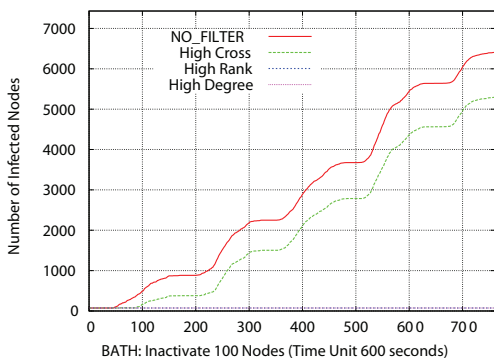

Figure 6: BATH: Impact of Hub Nodes in Epidemic Spread

5. PARTY AND DATE HUBS

When two hub nodes have the same degree distribution in time-dependent networks, they could have same degree distributions at each time unit, or each node might have different degree distributions in time and or space. Fig. 12 shows two distinct types of hub nodes called *party hubs* and *date hubs*, where *party hubs* always connects to the same nodes while *date hubs* connect to different nodes in time. Both could show high degree distributions, but the function as a hub is different. If we can distinguish these two types of hubs, it will be useful for controlling the information flow. Thus, we investigate the existence of two distinct types of hubs: *party hubs* and *date hubs* in a pair interaction in human connectivity traces.

In a biological study, Han et al. [9] describe that hubs in the protein to protein interaction network of *yeast* have been classified as *party* hubs, which are highly correlated in their *mRNA* expression with their partners during the cell cycle while *date* hubs show less correlation. The study shows that intrinsic disorder is significant in date hub proteins when compared with party hub proteins. The enrichment of intrinsic disorder in date hubs may facilitate transient interactions, which might be required for date hubs to interact with different partners at different times. Fewer party hubs contain long disordered regions compared to date hubs, indicating that these regions are essential for flexible binding, while party hubs interact to a large number of other proteins, supporting the idea of party hubs as the cores of highly clustered functional modules.

Removal of date hubs may split the network into many fragments, indicating a unique role of date hubs in global network topology and, possibly, resilience to genetic disturbances. Because sub-network fragments formed upon deletion of date hubs are functionally more homogeneous than fragments formed upon deletion of party hubs, it appears that party hubs reside within single modules that perform biologically discrete tasks, whereas date hubs mediate communication between different modules.

Cell cycle is molecular operation in cells, and complex human networks may be explained with similar principles. In the next sections, we explore human connectivity traces to examine if party and date hubs exist.

5.1 Neighbourhood Similarity Rate

We define an intuitive formulation of the *Neighbourhood Similarity Rate (NSR)*. It indicates the dynamic change of neighbourhood nodes, therefore a larger NSR confirms that connectivity does not change much. A lower NSR indicates that neighbours enter and leave a neighbour set.

$$NSR = \ N_t \cap N_{(t+1)}/N_t \cup N_{(t+1)} \tag{1}$$

where N is a set of neighbour nodes and t for *timeunit*.

The NSR could be used as a measure to distinguish *party* and *date hubs*, where a higher NSR value indicates *date hub* and a lower value indicates *party hub*. We have calculated NSR values of *DEGREE* hubs described in Section 4. To process the sparse traces and highlight the change of neighbourhood node sets, we suppress the node information at the timeunit when no connectivity is recorded. Thus, $t + 1$ in (1) indicates the next timeunit when the node has a connection with at least one node in this paper. Fig. 7 depicts the NSR of MIT trace. The timeline in Fig.7-11 indicates one day duration of node activities. Fig. 7 a) shows the characteristics of the party hub, where nodes are together most of time, while Fig. 7 b) and c) depict date hub characteristics. For the date hubs, the figures have a limit, only showing the evolution of rate but not the size of neighbourhood. Fig. 7 d) shows the actual neighbourhood nodes among 97 nodes. The change of NSR maps clear change of
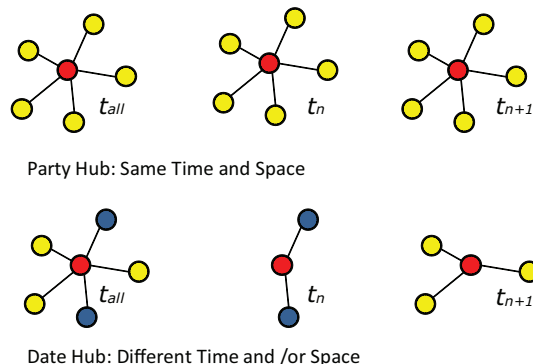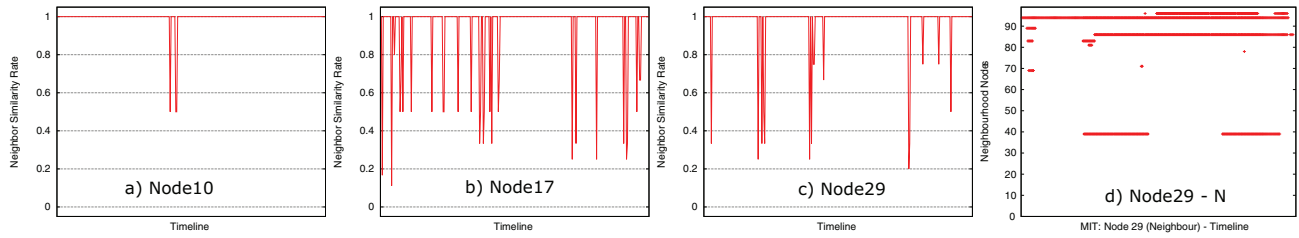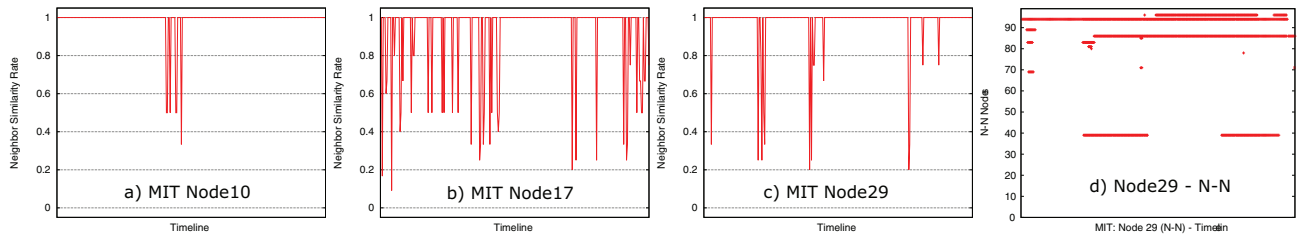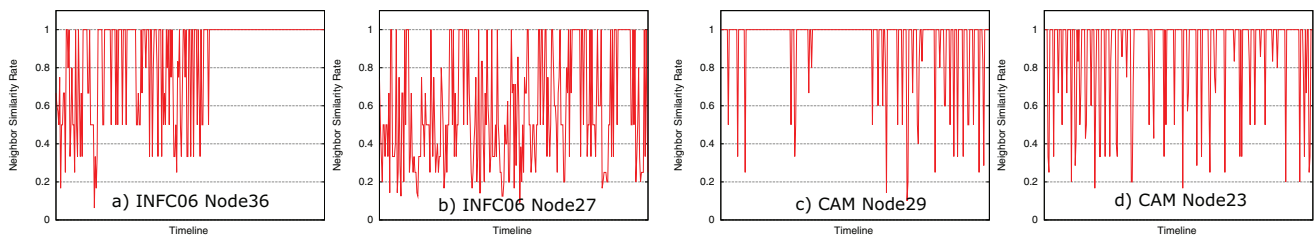

Party Hub: Same Time and Space

Date Hub: Different Time and /or Space

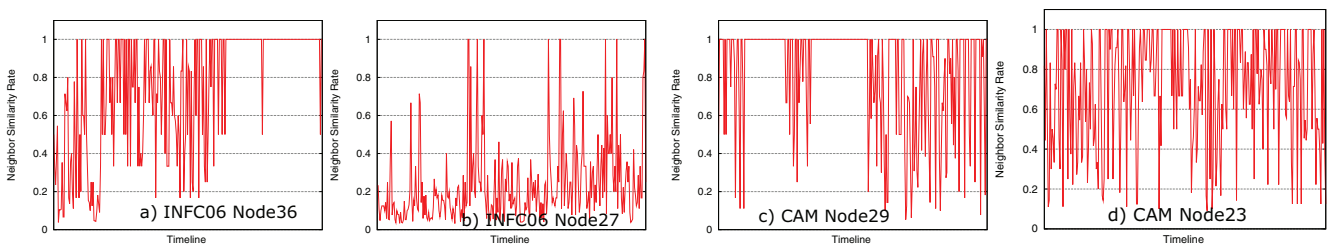**Figure 12: Party Hub and Date Hub**

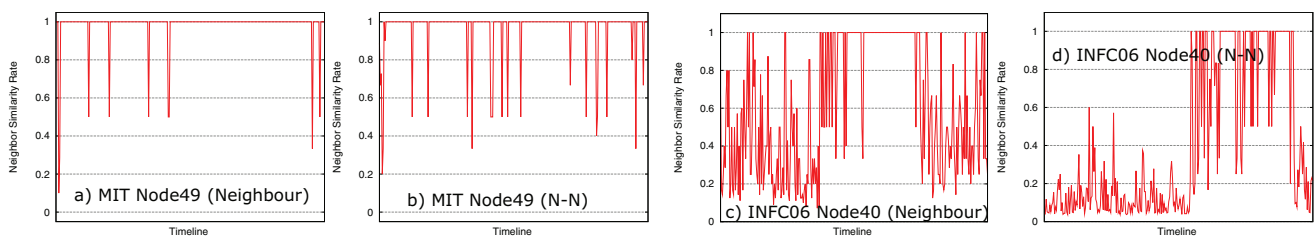**Figure 7: MIT: Neighbourhood Similarity Rate**

Figure 8: MIT: Neighbourhood plus Neighbourhood Similarity Rate

Figure 9: INFC06/CAM: Neighbourhood Similarity Rate

Figure 10: INFC06/CAM: Neighbourhood plus Neighbourhood Similarity Rate

Figure 11: Different Hub Types (Neighbourhood and Neighbour-Neighbourhood)

the neighbourhood node set. Fig. 9 a) and b) depict the NSR of the INFC06 trace and Fig. 9 c) and d) show the NSR of the CAM trace. We calculated NSR of top 30 DEGREE Hubs. On average, 30% of hub nodes show the characteristics of party hub, 40% for date hub, while the rest shows the combined form. The combined case displays time dependent hub characteristics. The BATH data results always in NSR value 0, which reveals extremely sparse network in the urban scale trace.

Fig. 8 and Fig. 10 depict the same calculation using the *Neighbourhood plus Neighbourhood Similarity Rate (NNSR)*, where NN is calculated in the same way as NSR, except using a set of neighbour nodes including all the sets of neighbour nodes of each neighbour node. The NSR and NNSR show same characteristics in Fig. 8 and Fig. 10. On the other hand, Fig. 11 a) and b) depict slightly different characteristics, where Fig. 11 a) shows distinct *party hub* characteristics but not as significant as in Fig. 11 b). Fig. 11 c) and d) depict a clear difference between the NSR and NNSR. The cause of the change could be fast node movement, and further investigation is in progress.

In Section 4, we have shown a high correlation and membership similarity between RANK Hub and DEGREE Hub. RANK Hubs may split into two types of party and date hubs, where the former contributes to relay messages towards the inside of the cluster, while the latter takes a task to link between clusters. Both types of hubs exhibit a high degree distribution. Locally observing NSR at the node helps to identify node ranking, making a distributed operation possible instead of calculating ranking in centralised manner.

6. RELATED WORK

The recent discovery of the properties of complex networks in a structure of biological and social systems [18] brought different perspectives on real world networks. Traditionally random networks have been extensively studied [6]. Random graphs are usually constructed by randomly adding links to a static set of nodes and tend to have short paths between pair nodes. Power-law networks are networks where the probability that a node has the degree k is proportional to k. Many real-world networks have been shown as power-law networks, including Internet topologies and social networks.

Emerging wireless technologies are creating a physical network in the actual physical space along virtual communication. 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] allows us to see the insight of social networks.

7. CONCLUSION AND FUTURE WORK

In this paper, we have presented our study of characteristics of hubs in dynamic human networks from human connectivity traces. Hubs can de-fragment the network and contain distinct characteristics to influence data flow within the network. We are currently working on further analysis on party and date hubs with our previous work on community detection. The pervasive DTNs are dynamic, and we are particularly interested in how the network structure affects the information flow, and vice versa, how the ongoing communication affects the network structure. Pair communication and social structure need to be integrated and modelled along dynamic interaction. The social network reflects access to information, and change of social activities can be seen as seeking better information access. Our ultimate goal is complete understanding of human-to-human network models in the urban space.

8. REFERENCES

[1] Albert, R. and Barabasi, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47, 2002.

[2] Chaintreau, A. et al. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proc. INFOCOM*. 2006.

[3] College, D. A community resource for archiving wireless data at dartmouth, http://crawdad.cs.dartmouth.edu/index.php, 2007.

[4] Haggle Project, http://www.haggleproject.org, 2008.

[5] Eagle, N. and Pentland, A. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006.

[6] Erdos, P. and Renyi, A. On random graphs i. *Mathematicae*, 5, 1959.

[7] Fall, K. A delay-tolerant network architecture for challenged internets. In *Proc. SIGCOMM*. 2003.

[8] Freeman, L. C. A set of measuring centrality based on betweenness. *Sociometry*, 40:35–41, 1977.

[9] Han, J.-D. J., Bertin, N., Hao, T., et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430, 2004.

[10] Henderson, T. et al. The changing usage of a mature campus-wide wireless network. In *Proc. Mobicom*. 2004.

[11] Hui, P., Crowcroft, J., and Yoneki, E. BUBBLE Rap: Social Based Forwarding in Delay Tolerant Networks. In *proc. MobiHoc*. 2008.

[12] Hui, P., E.Yoneki, et al. Distributed community detection in delay tolerant networks. In *Proc. MobiArch*. 2007.

[13] Kleinberg, J. The wireless epidemic. *Nature*, 449(20), 2007.

[14] Leguay, J. et al. Opportunistic content distribution in an urban setting. In *ACM CHANTS*. 2006.

[15] Newman, M. Analysis of weighted networks. *Physical Review E*, 70:056131, 2004.

[16] Nicolai, T., Yoneki, E., et al. Exploring social context with the wireless rope. In *Proc. MONET: LNCS 4277*. 2006.

[17] O'Neill, E. et al. Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In *UbiComp*, pages 315–332. 2006.

[18] Strogatz, S. H. Exploring complex networks. *Nature*, 410 (268-276), 2001.

[19] UCSD. Wireless topology discovery project, http://sysnet.ucsd.edu/wtd/wtd.html, 2004.

[20] Winters, P. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6:324–342, 1960.

[21] Yoneki, E., Hui, P., et al. A socio-aware overlay for multi-point asynchronous communication in delay tolerant networks. In *Proc. MSWiM*. 2007.

[22] Yoneki, E., Hui, P., and Crowcroft, J. Visualizing Community Detection in Opportunistic Networks. In *ACM CHANTS*. 2007.

[23] Yoneki, E., Hui, P., and Crowcroft, J. Epidemic Spread in Dynamic Human Networks. In *Bio-Inspired Computing and Communication*. 2008.