Visualizing Community Detection in Opportunistic Networks

Eiko Yoneki, Pan Hui, Jon Crowcroft
University of Cambridge, Computer Laboratory
Cambridge CB3 0FD United Kingdom
[firstname.lastname@cl.cam.ac.uk]

ABSTRACT

Community is an important attribute of Pocket Switched Networks (PSNs), since mobile devices are carried by people who tend to belong to communities in their social life. We discover the heterogeneity of human interactions such as community formation from real world human mobility traces. We have introduced novel distributed community detection approaches and evaluated with those traces [11]. This paper describes a series of visualizations to show characteristics of human mobility traces including community detection. We focus on extracting information related to levels of clustering, network transitivity, and strong community structure. The progression of the connection map along the community formation process is also visualized.

Categories and Subject Descriptors

C.2.4 [**Computer Systems Organization**]: Computer Communication Networks—*Distributed Systems*; I.6 [**Computing Methodologies**]: Simulation and Modeling

General Terms

Measurement, Experimentation, Algorithms

Keywords

Distributed Community Detection, Delay Tolerant Networks, Network Measurement, Social Networks

1. INTRODUCTION

The Haggle project [6] introduced the Pocket Switched Network (PSN) [2], a type of Delay Tolerant Network (DTN) [8] that provides intermittent communication for humans carrying mobile devices. It is essential to understand human behavior to build PSNs. Thus, a number of experiments to capture the human mobility in the real world have been conducted. For example, in the Haggle and Wireless Rope projects [1], human mobility and proximity with Bluetooth have been studied. The characteristics of these data, such as inter-contact and contact distribution, have been explored in several studies [2], to which we refer the reader for further background information. These studies demonstrate that mobility gives rise to local connection opportunities when access infrastructure is not available.

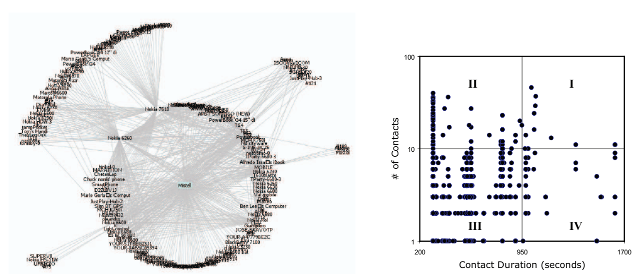
The collected device information in the Wireless Rope project [14] is visualized in real-time on a web (e.g. degree separation in Fig.1a). Users can explore their own neighbourhood including contacts, regularly met familiar strangers and randomly encountered strangers. Node pair characteristics are shown in four categories in Fig. 1b (see Section 3.1 for more detail). The connection map is a tool for personal social network analysis, e.g. to identify common contacts and distinct cliques.

To improve understanding of DTNs, we focus on community detection from the mobility traces. We identified the uniqueness of DTNs in comparison to the generic weighted networks studied by Newman [12]. We adapted this weighted network analysis to DTNs and evaluated our community detection algorithms over real-life human mobility traces (more detail in [9] [11]). Furthermore, we introduced distributed community detection [11], which can be deployed in real-time by distributed applications embedded in mobile devices.

This paper presents a snapshot of our ongoing work on a series of visualizations to show characteristics of mobility traces including community detection. We focus on extracting information related to levels of clustering, network transitivity and strong community structure. Particularly, we visualize the detected communities from the mobility traces as well as progression of the connection map along the community formation process. Different community detection algorithms can be applied. The analyzed traces include the data from the MIT Reality Mining project [7], the UCSD wireless topology discovery project [16] from the Crawdad database and the Haggle project [6]. The Wireless Rope project collected conference activity data [14]. See [11] for further details of each trace data. The visualization tool is generic and any trace data following the the required format can be visualized.

2. COMMUNITY DETECTION

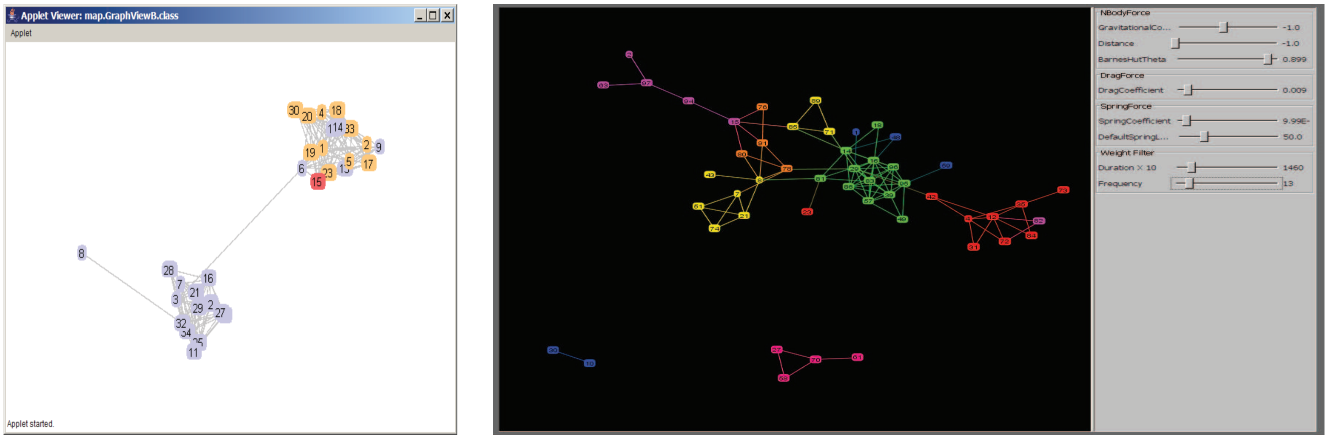
Community is an important attribute, because mobile devices are carried by people who normally belong to communities in their social life. If the correlated interaction concept applies, then our intuition is to use this community information to influence information dissemination mechanism including use of various centralities intra/inter communities. Hence identifying local communities of each mobile device can be important to improve data forward-



a. Connection Map b. Node Classification

Figure 1: Device Connection Visualization

CHANTS'07, September 14, 2007, Montréal, Québec, Canada.
ACM 978-1-59593-737-7/07/0009.



a.Cambridge iMote Trace **b. MIT Trace**

Figure 2: Detected Communities

ing efficiency. We have shown significant impact of community detected from mobility traces on forwarding algorithms [10][9].

Many centralized community detection methods have been proposed and examined in the literature (see recent review papers by Newman [13] and Danon *et al* [5]). These centralized methods are useful for offline data analysis on mobility traces collected to explore structures in the data. But, as they form self-organizing networks, we would also ask whether the mobile devices can sense and detect their own local communities instead of relying on a centralized server, which leads to the area of distributed community detection. Clauset [3] defines a measure of local community structure and an algorithm that infers the hierarchy of communities by enclosing a given vertex by exploring the graph one vertex at a time.

We proposed and evaluated several novel distributed community detection approaches with great potential to detect both static and temporal communities [11]. Two of our detection algorithms are SIMPLE and k-CLIQUE, where SIMPLE is based on [3] by classifying the contact duration of a node pair according to an *a priori* threshold value. k-CLIQUE is based on [15] in which a k-clique community is defined as the union of all k-cliques (complete subgraphs of size k) that can reach each other through a series of adjacent k-cliques, where two k-cliques are said to be adjacent if they share $k-1$ nodes. Details of the evaluation methodologies and

Experimental data set	SIMPLE	k-CLIQUE
Reality	0.79/0.76	0.87
UCSD	0.47/0.56	0.55
Cambridge	0.85/0.85	0.85

Table 1: Summary of Distributed Community Detection

results can be found in [11]. Table 2 summarizes the highest similarity values calculated by each distributed algorithm. For SIMPLE, we show both its comparison with the centralized k-CLIQUE (first) and the centralized Newman method [12] (second). k-CLIQUE has slightly better performance than its SIMPLE counterpart, because k-CLIQUE requires more information and calculation. Considering its computational and storage requirements, the performance of SIMPLE is acceptable. The complexity of SIMPLE is $O(n)$, and it may be suitable for resource constrained mobile devices. If the mobile devices can afford more storage, k-CLIQUE would be a good choice due to its reasonably good similarity values.

3. VISUALIZATION

We develop visualization tools ranging from device connection maps to distributed community detection by replaying mobility trace in a discrete event emulator. Selected examples are shown below, but these do not comprise on exhaustive list. In PSNs, data dissemination may be navigated by human interaction, which may be influenced by human observation of the network state. Thus, visualization of the network state is particularly important.

3.1 Four Node Categories

The correlation between contact duration and the number of contacts can be split into the four categories below. Meetings take place between pairs of individuals at a rate which is high if a pair has one or more mutual friends and low otherwise. Acquaintances are between pairs of individuals who rarely meet and decay over time. There is an upper limit for the number of friendships an individual can maintain. Proximity determines community in many cases; however, how to evaluate proximity or common interests is an issue still to be determined. A classification chart with selected or individual pairs can be shown. We experiment with various threshold values for our community detection. Fig.1b depicts the four categories on a Wireless Rope trace.

I Community High $N°$ of contacts and long contact duration

II Familiar Stranger High $N°$ of contacts and short contact duration

III Stranger Low $N°$ of contacts and short contact duration

IV Friend Low $N°$ of contacts and long contact duration

3.2 Detected Communities

Fig. 1 depicts connection maps after community detection using k-CLIQUE. In Fig. 1a, two distinct communities are visible, where undergraduate students from two different classes participate in the experiment. Fig. 1b depicts a larger experiment, with around 100 devices. Fig. 1b is the result of running the first 3 months trace among 9 months of total duration. After running for the whole 9 months, eight communities are recognized. The four communities are clearly recognized in Fig. 1b, while the others are not yet visible. Two sliding buttons can be used to change the threshold value on the duration of contact time and the number of contacts, which should show different community formation. Furthermore, different algorithms of community definition can be applied, which would show different connection maps.

3.3 Evolution of Connection Map

We also visualize the evolution of a connection map by replaying the mobility trace in a discrete event emulator. Fig. 3.1 illustrates a snapshot of progression. The time filter provides the function

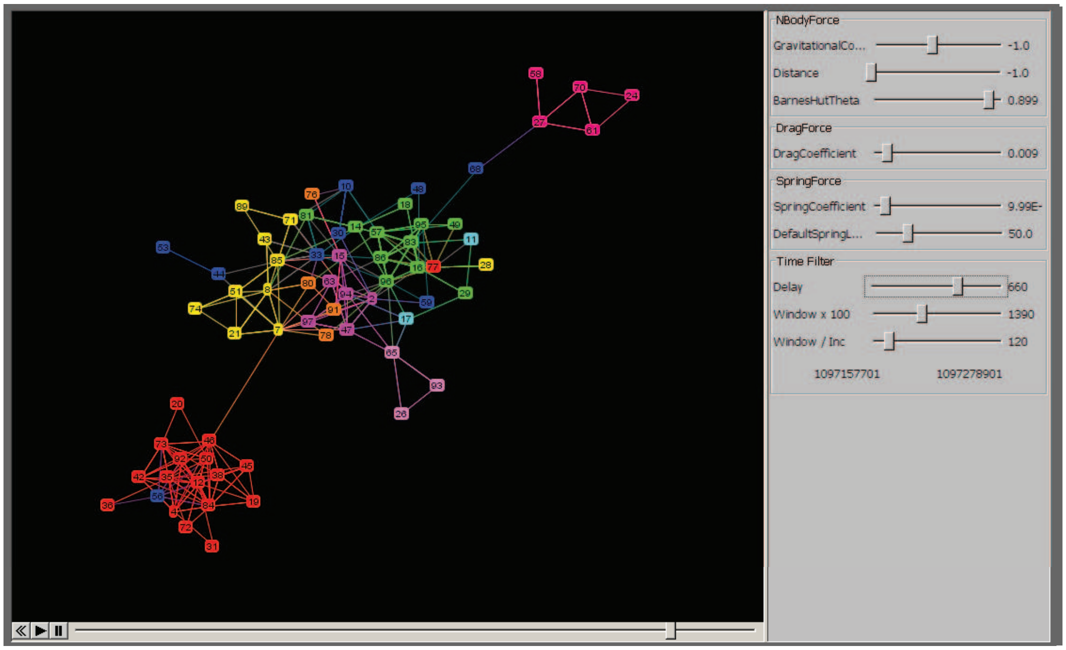


Figure 3: Progression of Connection Map with Time Filtering (MIT Trace)

to change the *delay*, *window size*, and *window increment*, which shows the topology of communities dependent on the time window. Fig. 3.3 and Fig. 3.3 depict several snapshots of topology change along the timeline, which changes the relation between pair nodes by calculation of community detection. The darker edge color indicates a higher contact duration and the number of contacts based on the threshold value.

Fig. 3.3 shows the same operation, and the nodes are colored with their respective communities found by global detection. Fig. 3.3 depicts the connection map after running the whole trace, where the combined view includes all edges, and the community view shows only edges within the community. The speed of the discrete event emulator can be changed and paused so that a specific time period can be observed more precisely. Different community detection algorithms can be applied.

4. CONCLUSION

Visualization of human mobility traces highlights community detection and device interaction, and provides rich insight into social behavior. Communities and societies contain structures for social networking, and these structures can be powerful for exploiting the information flow. We investigated how the local and global characteristics of the network can be used practically for information dissemination. As a next step, we are working on incorporating spatial information and device correlation patterns in visualization.

Acknowledgement. This research is funded in part by the Haggle project under the EU grant IST-4-027918. We would like to acknowledge Can Altineller for his help on implementing the visualization tool. We would also like to acknowledge CRAWDAD project [4] for their hosting and sharing of the mobility data.

5. REFERENCES

[1] Wireless Rope, http://sourceforge.net/projects/wirelessrope, 2006.

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

[3] A. Clauset. Finding local community structure in networks. *Physical Review E*, 72:026132, 2005.

[4] D. College. A community resource for archiving wireless data at dartmouth, http://crawdad.cs.dartmouth.edu/index.php, 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, 2004.

[7] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, V10(4):255–268, May 2006.

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

[9] P. Hui and J. Crowcroft. Bubble rap: Forwarding in small world dtns in every decreasing circles. Technical Report UCAM-CL-TR684, Univ. of Cambridge, 2007.

[10] P. Hui and J. Crowcroft. How small lables create big improvements. In *Proc. IEEE ICMAN*, March 2007.

[11] P. Hui, E.Yoneki, S. Chan, and J. Crowcroft. Distributed community detection in delay tolerant networks. In *Proc. MobiArch*, 2007.

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

[13] M. Newman. Detecting community structure in networks. *Eur. Phys. J. B*, 38:321–330, 2004.

[14] T. Nicolai, E. Yoneki, N. Behrens, and H. Kenn. Exploring social context with the wireless rope. In *Proc. Workshop MONET: LNCS 4277*, 2006.

[15] G. Palla et al. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.

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

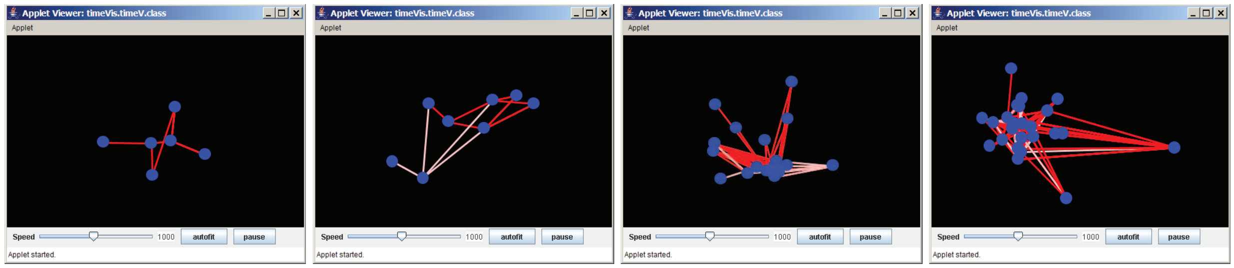


Figure 4: Evolution of Connection map and Edge Characteristics (MIT Trace)

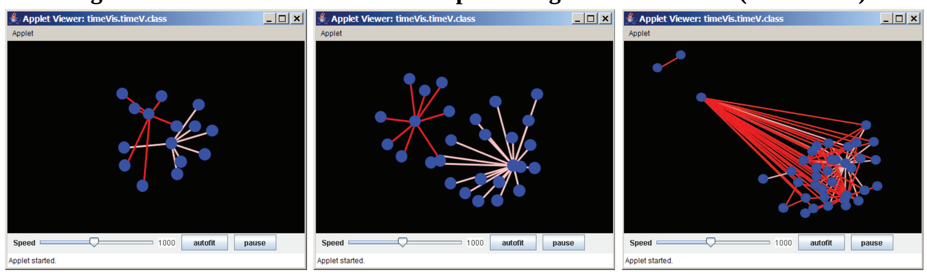


Figure 5: Evolution of Connection Map and Edge Characteristics (UCSD Trace)

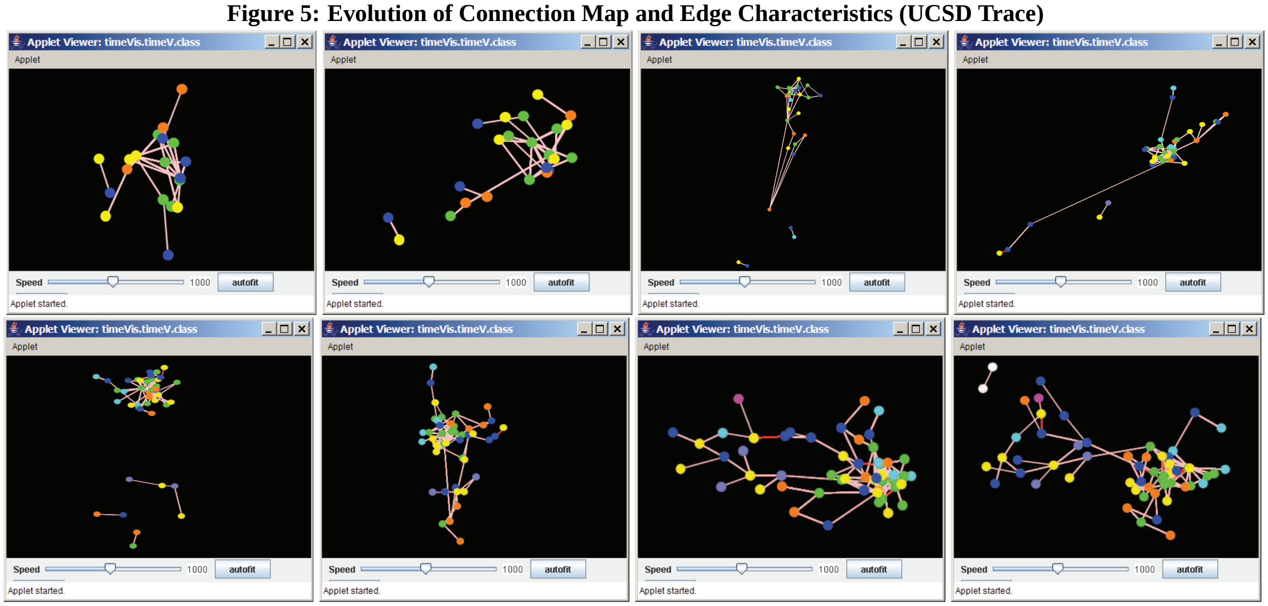
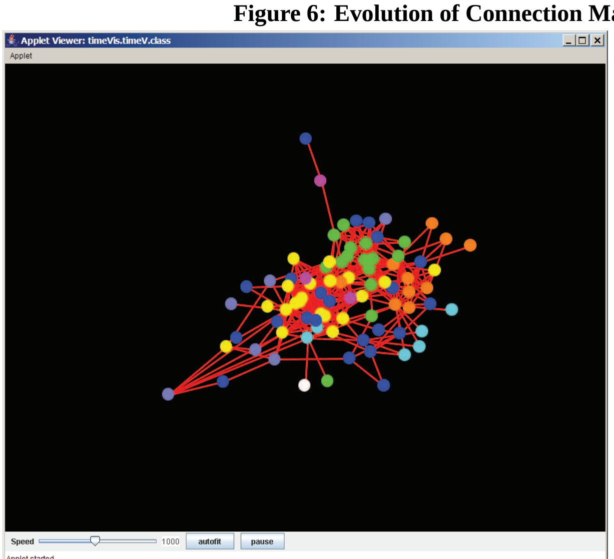
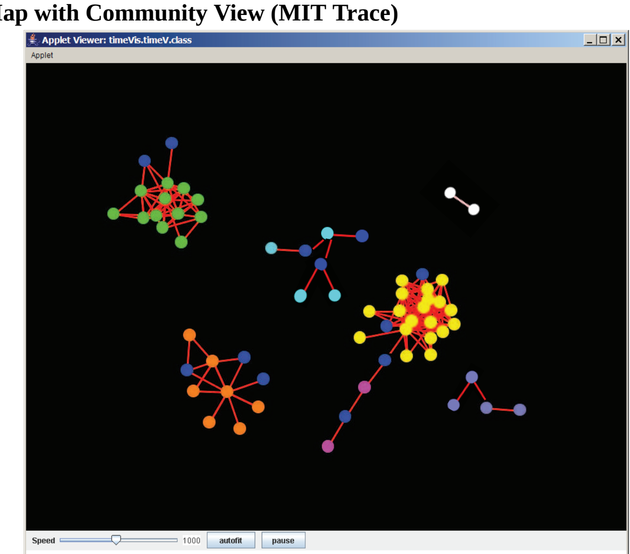


Figure 6: Evolution of Connection Map with Community View (MIT Trace)



a. Combined View



b. Community Separation View

Figure 7: Evolution of Connection Map (Duration 9 Months - MIT Trace)