

Dynamics of Inter-Meeting Time in Human Contact Networks

Eiko Yoneki, Dan Greenfield, and Jon Crowcroft

Computer Laboratory

University of Cambridge

Cambridge, United Kingdom

{eiko.yoneki, dan.greenfield, jon.crowcroft}@cl.cam.ac.uk

Abstract—We envision new communication paradigms, using physical dynamic interconnectedness among people. Delay Tolerant Networks (DTNs) 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). Information propagation in PSNs is highly influenced by human connectivity networks, i.e. social networks.

In our previous work, we have exploited constructing weighted networks using characteristics of pair connections such as the duration of contact time and frequency of contacts from time series of human connectivity network traces. The approach we took is based on empirical and heuristic and the focus is finding a single aggregated logical network structure. The physical network topology in the real world is time dependent and it is a complex task to describe its dynamics. This paper aims to identify dynamics of meeting groups in human connectivity traces, where meeting groups are expected to be a group interacting among the nodes in physical space. Thus, we define ‘meeting group’ differently from ‘community’. We exploit statistical approach that provides quantitative attributes to uncover meeting groups. We identify the power law behavior of meetings that is important for supporting to understanding dynamics of information flow between meeting groups and building group oriented communication protocol.

Keywords-complex networks; K-Clique; inter-meeting time

I. 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 among people, leading towards a world where digital traffic flows as people pass each other [13]. Delay Tolerant Networks (DTNs) [8] 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) [3].

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, decentralized and fast. However, careful tuning to achieve reliability and minimize network load is essential. 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.

We have previously reported an approach that uses a logical connection topology, and that uncovers hidden stable network structures, such as social networks [12] [23], from the human connectivity traces. We have shown improved performance by applying these extracted social contexts to a controlled epidemic strategy [11]. During this work, we have realized 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.

A series of studies has shown the statistical properties of social networks such as the small-world and power-law degree distributions. For community detection, various methods have been reported. Clustering of networks, where network nodes sometime form cliques or loose connections from time to time is a significant property in human movement. In our previous work, we have exploited community detection from the human connectivity traces by constructing weighted networks using characteristics of pair connections such as the duration of contact time and frequency of contacts also exploited spectral properties of the graph as well as Laplacian matrix [22]. Mostly, the approach we took is based on empirical and heuristic and the focus is finding a single aggregated logical network structure called ‘community’. For dynamic graph mining, Berger-Wolf et. al. show the study of community evolution based on node overlapping [2]. The evolution of subgraphs over time in biological networks has been discussed, however, these studies are based on static network setting and fairly small scale. The traces described in the next section are more complex with thousands of updates per day.

We define ‘meeting group’ differently from ‘community’. Actual meeting among the member of the community may occur at certain time or location for possibly predictable duration. The number of participating members may not be 100% of the community members. Thus, it is important that the concept of community differs from ‘meeting group’. Meeting groups can be the base of inferring the community.

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

Table I
Characteristics of the experiments

Tracking the dynamics of the meeting should show the inter-relationship of members within the community.

Our goal is inferring dynamics of meetings in human connectivity networks based on the traces collected human connectivity by sensors. Our empirical data based approach will give real insight of time-dependent dynamic network modeling. However, we encounter several problems to infer ‘meeting group’ from the human connectivity traces such as too many missing edges between nodes.

Thus, we claim our contribution in this paper is two-fold: 1) sharing the issues we encounter for inferring dynamics of meeting groups 2) as preliminary result, we show the power law behavior of meetings and the significance of meeting groups using simpler methods. This demonstrates duration of meetings and inter-meetings for predicting network capacity or the limit of synchronisation mechanism. Even with noisy data, we believe that the result can lead to understanding dynamics of information flow between meeting groups. The rest of this paper is structured as follows. We introduce the experimental data sets in Section 2, and then describe the proximity detection mechanism by Bluetooth communication in Section 3. In Section 4, we analyse the duration of meetings and inter-meetings followed by brief discussion of related works. Finally, we conclude the paper with a brief discussion and future works in Section 6.

II. 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 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] [19] use Bluetooth to measure device connectivity, while others [10] 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 [4] listed below, and Table I 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 [21], approximately 300 wireless PDAs running Windows CE were used to collect WiFi access point information periodically for 11 weeks.

CAM: in the Cambridge Hagggle project [14], 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 [3].

III. PROXIMITY DETECTION WITH BLUETOOTH

Bluetooth is a low-power open standard for Personal Area Networks (PANs) and has gained its popularity due to its emphasis on short-range, low-power and easy integration into devices. The platform used in the Hagggle experiments is the Intel Mote ISN100-BA (known as the iMote). The iMote runs TinyOS and is equipped with an ARM7TDMI processor operating at 12MHz, with 64kB of SRAM, 512kB of flash storage, and a multi-colored LED, and a Bluetooth 1.1 radio. The specifications lists the radio range to be 30 meters.

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 uses a special physical channel for devices to discover each other. A device becomes discoverable by entering the inquiry substate where it can respond to inquiry requests. The inquiry scan substate is used to discover other devices. The discovering device iterates (hops) through all possible inquiry scan physical channel frequencies in a pseudo-random fashion. For each frequency, it sends an inquiry request and listens for responses. Therefore, a Bluetooth device cannot scan for other devices and be discoverable at the same time. 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 Bluetooth standard [18], recommends being in the inquiry scan substate for 10.24 seconds in order to collect all responses in an error-free environment. The power consumption of Bluetooth also limits the scanning interval, if devices have limited recharging capability. The iMote connectivity traces in Hagggle use a scanning interval of approximately 2 minutes, while the Reality Mining project uses 5 minutes. The ratio of devices with Bluetooth enabled to the total number of devices is around only an average 15% of population.

Bluetooth for proximity detection is widely available and a lot of people carry a Bluetooth enabled mobile phone with them. Thus, it is possible to detect a certain amount of peoples phones without handing a special device to each of them, which makes Bluetooth appealing for experiments involving a large quantity of people. The range of Bluetooth

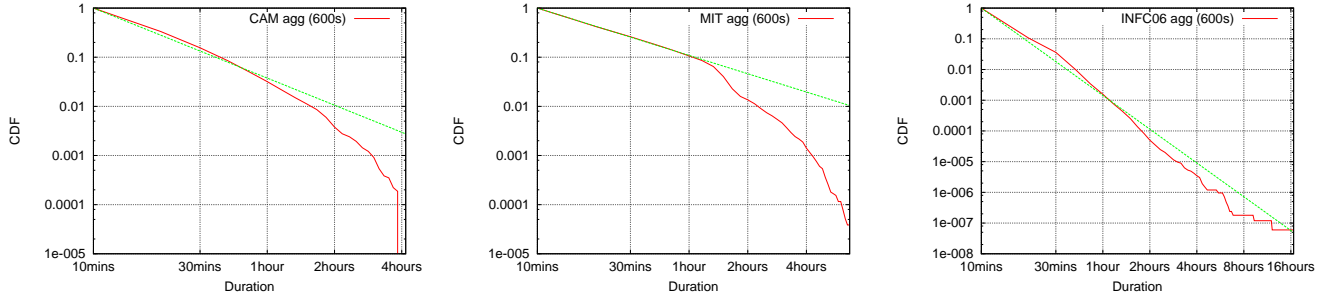


Figure 1. Distribution of Meeting Times - Aggregated all K results

varies between 10m and 100m, depending on the device class. In mobile phones, the range is usually 10m. We have observed the devices can be detected in 20m range if there is no obstacles, while if there is any obstacles such as a thick wall it limits to 5m range.

The logged data from the above experiments are used to reconstruct the time-dependent connectivity information to study the distribution of contact times, inter-contact times, and their statistical properties, where we constructed discrete event traces of pair interactions of 600 seconds interval. We have aggregated raw data within 600 seconds time units to avoid uncertainty of device detection from various Bluetooth based mechanism. For the MIT data, part of around 8 months is used.

IV. MEETING TIME AND INTER-MEETING TIME

In this section, we show results for the analysis of meeting dynamics and focus on meeting time and inter-meeting time statistics. We have used various community detection algorithms in our previous work [11] and found K-CLIQUE [20] shows stable results for different types of human contact traces. Thus, we demonstrate inferring physical group meetings using K-CLIQUE algorithm in this paper. The inter-contact time is the time interval between two contacts. Inter-contact time is the duration from when one contact finishes and the next one begins, it determines how often a communication is possible. Shorter inter-contact time means that the two people see each other quite often. If two people have short inter-contact time, then it means that we can wait for the next contact and send data directly. For longer inter-contact times, we need to rely on other media to transfer the data instead of waiting unreasonably long. The number of such contacts and the distribution of contact durations is an important factor in determining the capacity of opportunistic networks. It gives insight on how much data can be transferred at each opportunity. This concept of pair node relationship can apply on the relationship between meeting groups, which helps building group oriented communication protocols.

In [3], we have shown that the distribution of inter-contact times of pair nodes follows an approximate power law over a large time range, where most nodes have fairly short interval

time between subsequent connections. The nature of the distribution affects the choice of suitable forwarding algorithms to be used to maximize the successful transmission of messages in a bounded delay. This result demonstrated that opportunistic transmission schemes could be significantly improved by using limited redundant transmissions.

We envision potential applications over PSNs will be more group oriented such as sharing information among the social groups or location based groups. Thus, understanding the dynamics of groups such as the duration of meetings called *meeting time* and the interval between meetings called *inter-meeting time* are important values, which address the following questions:

- When will a meeting occur again?
- How likely is it that there is sufficient time for sharing information or transmitting data among a meeting group?

A. Inferring Meetings by K-CLIQUE

Palla et al. define a k-clique community as a union of all k-cliques (complete subgraphs of size k) that can be reached from each other through a series of adjacent k-cliques [20]. Two k-cliques are said to be adjacent if they share k - 1 nodes. This definition is based on their observation that an essential feature of a community is that its members can be reached through well-connected subsets of nodes, and that there could be other parts of the whole network that are not reachable from a particular k-clique, but they potentially contain further k-clique communities.

A problem with using k-cliques for detecting meetings is that of uniqueness. A cluster of nodes might have more than one overlapping k-clique embedded in it. For example, the largest k-clique size in a cluster might be five, but there can still be two overlapping 5-cliques, and by definition, each of those 5-cliques has embedded within it 5 overlapping 4-cliques and so on. Thus one must not only be careful about counting only one of the two 5-cliques, but also not to count the 4-cliques and 3-cliques that are subsets as well.

The figures 1-3 demonstrate how the overlapping problem affects the statistical analysis. In Figure Fig.1 shows the CDF distribution of meeting times, where detected meetings from K=3 to K=max are aggregated. These thus include overlaps

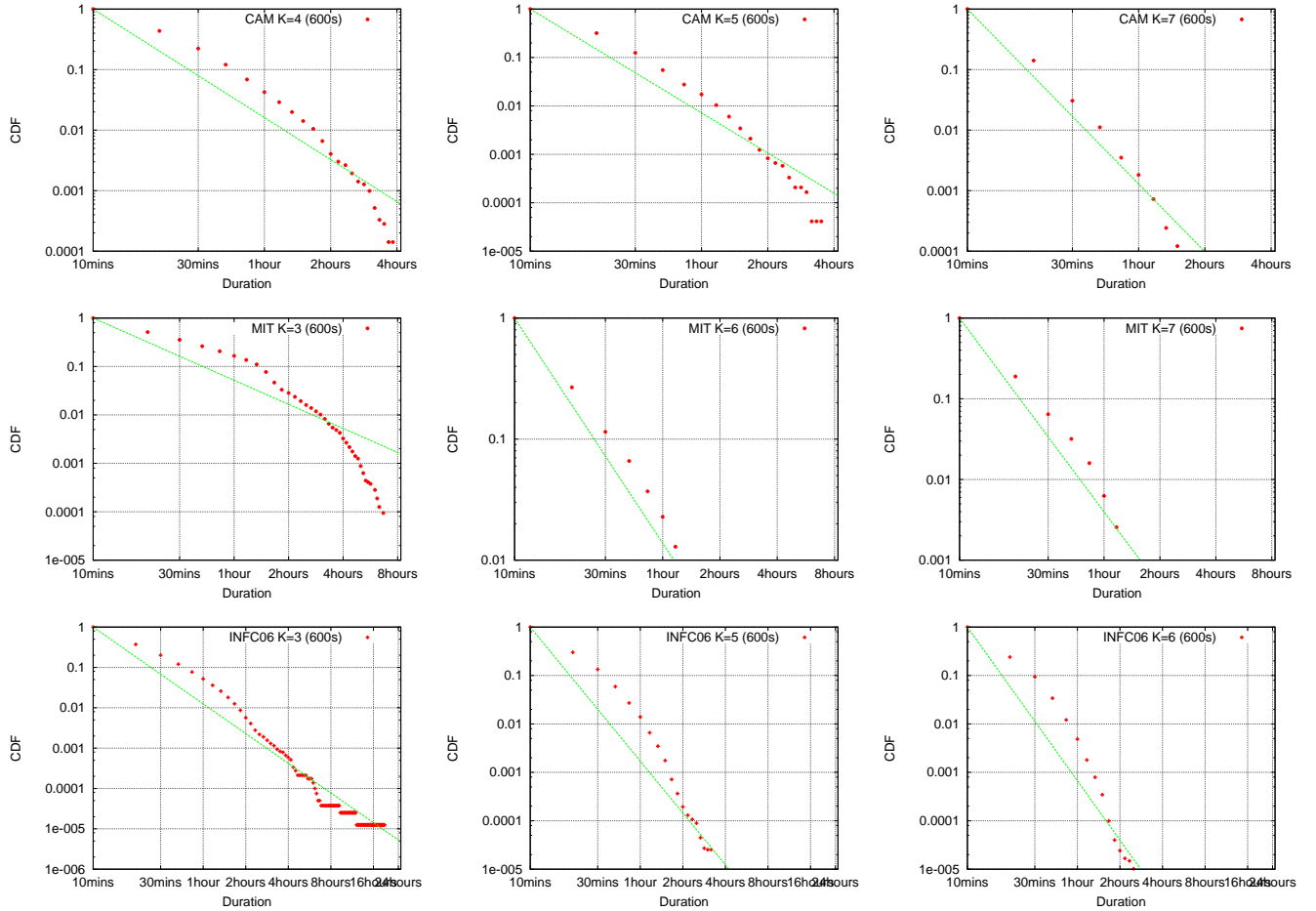


Figure 2. Distribution of Meeting Time - Individual K results

of each clique, thus a 5-clique would also be counted five times as a 4-clique, and so on. The distribution is then for the sum of these K-clique counts.

In more detail, Figure Fig.2 shows the individual distribution with all clique of size k including overlaps. We note that larger cliques exhibit more power-law behaviour in the MIT dataset, whilst for the INFC06 dataset it is more power-law in behaviour for lower clique sizes.

Fig.3 shows the distribution, where overlapping cliques are removed. We try to remove overlapping sub-cliques by working backwards from a high-clique base and removing all sub-cliques from the results. For example, when the base $K=7$, in CAM data, all subgraphs in $K=6$ to $K=3$ are removed for the duration of a 7-clique meeting. This ensures that cliques of size k are only counted once as size k. The figure shows a more clear power law distribution with a reduced rms error in the fit. We note however that, in the presence of noise, removing subgraphs is not a straightforward process without knowing the detection failure rate. For example, nodes a-b-c may form a 3-clique from times 1-10, then a 4-clique a-b-c-d from times 11-15, then b-c-d from times 10-20. The clique elimination used here was to count higher cliques as having precedence. Thus

the duration of the cliques in this example detects times 11-15 as a 4-clique, and separately detects times 1-10 and 15-20 as 3-cliques.

B. Discussion

Fig.4 shows the distribution of the inter-meeting time for several K values in K-Clique based meeting detection. We immediately note that the bulk of inter-meetings times are within 24 hrs. However, the distribution does not appear very power-law in its nature, except perhaps for the early head of the distribution. While the interaction times between nodes is not power-law, the duration of meetings does appear to be. We can explain this discrepancy by noting that meetings involve many nodes and that counting these as pairs of interactions leads to weighting the meeting duration by the (often large) number of pairs, thus skewing the distribution. Furthermore, the iMote data is noisy, which can make estimating the contiguous duration of an interaction unreliable, with the probability of a break due to noise growing with the length of the meeting. This thus also skews pairwise interactions to be shorter than they actually are. By tracking the cluster instead, noise can be better tolerated thus more reliably estimating the meeting duration.

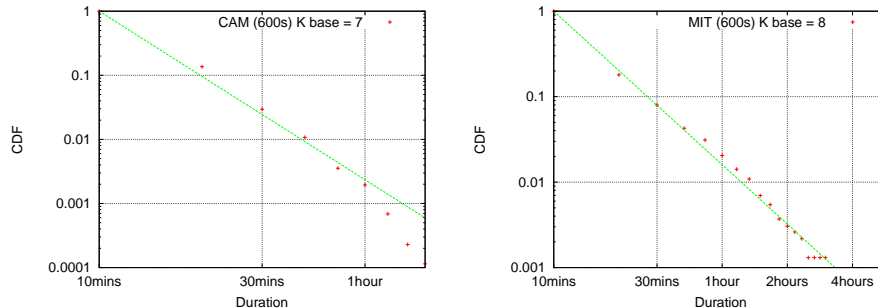


Figure 3. Distribution of Meeting Time without duplication

V. RELATED WORKS

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

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. 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 [16]. A weighted graph can be converted into a multi-graph with many unit edges. We only consider symmetric edges, on the other hand, edges can be symmetric (undirected) or asymmetric (directed), possibly with a different strength in either direction.

Community detection in complex networks has attracted a lot of attention in recent years. Community structures are usually substructures/subgraphs corresponding to important functions, and examples can be found in many areas, such as World Wide Web [9], biological social networks [17], and also the Internet [15]. The recent reviews [17] and [5] may serve as introductory reading in community detection methods. Besides the methods mentioned in the two reviews papers, we also introduce the *k*-CLIQUE community detection method by Palla *et al.* [20] and the weighted community analysis methods by Newman [16]. The main goal of these algorithms is to cluster nodes, while what we aim in this paper is inferring dynamic meeting group, which may eventually form communities. Our aim is also different from the concept of network evolution, where the evolution addresses the change of form. The work presented in this paper is a basic step to obtain clear understanding of dynamic human topology based on physical proximity.

VI. CONCLUSIONS AND FUTURE WORKS

We have shown the dynamics of meeting and inter-meeting time in this paper. The work presented in this paper has wide future extensions. Uncovering temporal and dynamics of meeting groups can be used as a signature for constructing synthetic network generation with the information of sub-graph structure and dynamics. Most importantly extracted model must be validated in some way of real experiments. The iteration of modelling and experiments will uncover further understanding of time-dependent complex human connectivity networks. Future works include: identifying the significance of meeting using transitivity, classifying behavior of nodes in the core/transient meetings, and dynamics of flow between meetings.

ACKNOWLEDGMENT

This research is funded in part by the EU grants for the Huggle project, IST-4-027918, and the SOCIALNETS project, 217141. We would like to acknowledge Derek Murray for helping the fast implementation of K-Clique algorithm and the CRAWDAD project [4] for their hosting and sharing of the connectivity/mobility data.

REFERENCES

- [1] R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47, 2002.
- [2] T. Berger-Wolf and J. Saia. A framework for analysis of dynamic social network. In *Proc. KDD*, 2006.
- [3] A. Chaintreau et al. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proc. INFOCOM*, April 2006.
- [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. Huggle Project, <http://www.huggleproject.org>, 2008.
- [7] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, V10(4):255–268, May 2006.

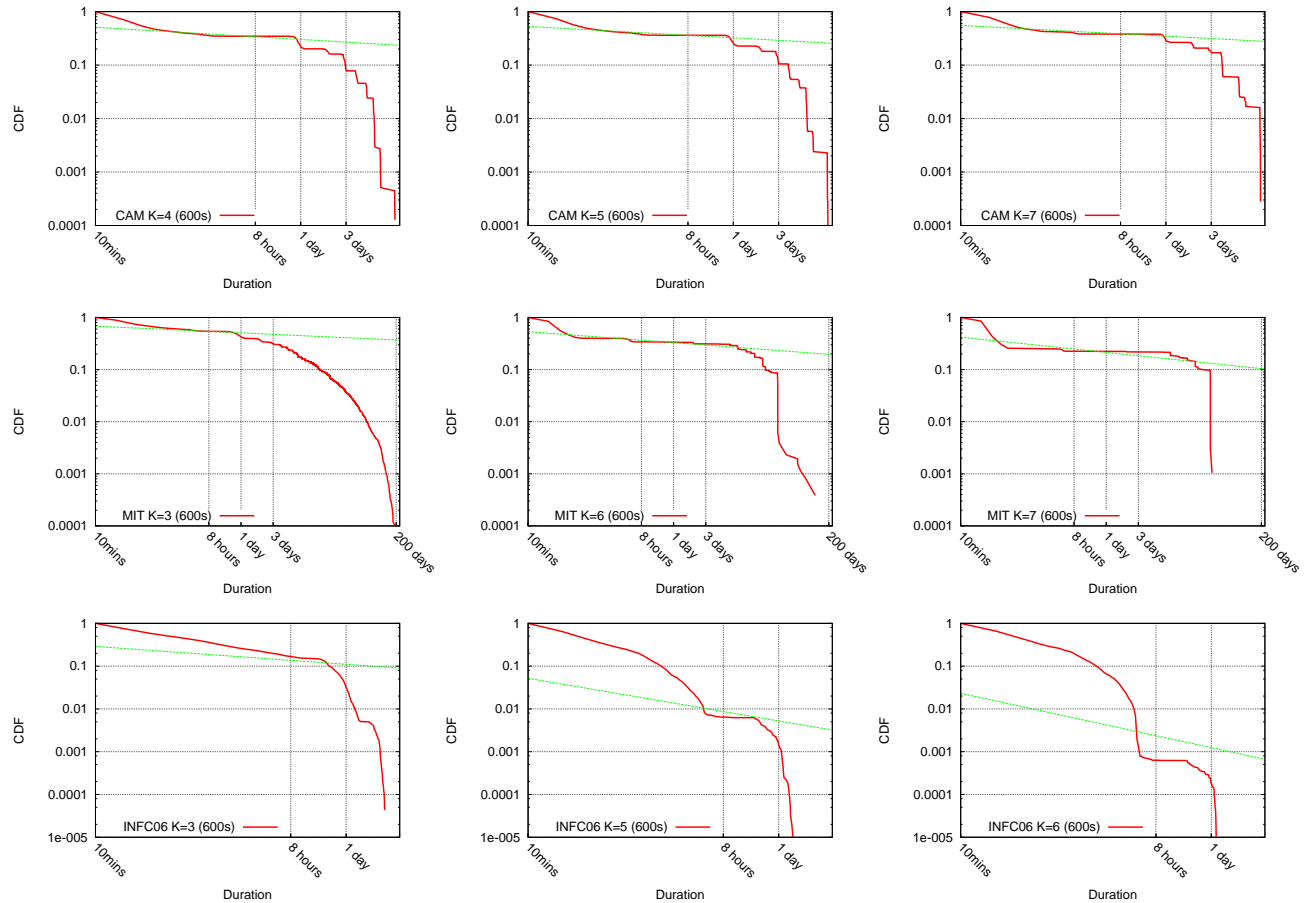


Figure 4. **Inter-Meeting INFC06**

- [8] K. Fall. A delay-tolerant network architecture for challenged internets. In *Proc. SIGCOMM*, 2003.
- [9] G. W. Flake, S. Lawrence, C. L. Giles, and F. Coetzee. Self-organization of the web and identification of communities. *IEEE Computer*, 35(3), 2002.
- [10] T. Henderson et al. The changing usage of a mature campus-wide wireless network. In *Proc. Mobicom*, 2004.
- [11] P. Hui, J. Crowcroft, and E. Yoneki. BUBBLE Rap: Social Based Forwarding in Delay Tolerant Networks. In *MobiHoc*, 2008.
- [12] P. Hui, E. Yoneki, S. Chan, and J. Crowcroft. Distributed community detection in delay tolerant networks. In *Proc. MobiArch*, 2007.
- [13] J. Kleinberg. The wireless epidemic. *Nature*, 449(20), 2007.
- [14] J. Leguay et al. Opportunistic content distribution in an urban setting. In *ACM CHANTS*, 2006.
- [15] D. Lusseau and M. E. J. Newman. Identifying the role that individual animals play in their social network. *PROC.R.SOC.LONDON B*, 271:S477, 2004.
- [16] M. Newman. Analysis of weighted networks. *Physical Review E*, 70:056131, 2004.
- [17] M. Newman. Detecting community structure in networks. *Eur. Phys. J. B*, 38:321–330, 2004.
- [18] T. Nicolai, N. Behrens, and E. Yoneki. Wireless rope: Experiment in social proximity sensing with bluetooth. In *IEEE Pervasive Computing and Communications (PerCom) - Interactive Experiment*, 2006.
- [19] T. Nicolai, E. Yoneki, N. Behrens, and H. Kenn. Exploring social context with the wireless rope. In *Proc. Workshop MONET: LNCS 4277*, 2006.
- [20] G. Palla et al. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.
- [21] UCSD. Wireless topology discovery project, <http://sysnet.ucsd.edu/wtd/wtd.html>, 2004.
- [22] E. Yoneki. Visualizing Communities and Centralities from Encounter Traces. In *ACM MobiCom - CHANTS*, 2008.
- [23] 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.