Findings from an Empirical Study of Fine-grained Human Social Contacts

Yi Wang
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 90089, USA
Email: wangyi@usc.edu

Bhaskar Krishnamachari
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 90089, USA
Email: bkrishna@usc.edu

Thomas W. Valente
Department of Preventive Medicine
University of Southern California
Los Angeles, CA 90089, USA
Email: tvalente@usc.edu

***Abstract*—An interaction based human contact study experiment has been conducted on 25 undergraduate students at USC, each carrying a wireless device (Tmote) for a week duration. Each mote transmits contact packets every 0.1 second to advertise its presence and a node receiving the packets will record the contact information. Data is processed off-line and a contact graph has been generated based on the strength of pairwise contact in order to visualize the grouping effect. All groups are identified and it has been found out that although most groups have small sizes and infrequent meetings, there exist large groups that have encountered several times in one week duration. The inter-contact and contact time distributions are found to be similar to findings from previous studies done in different settings. The inter-group contact time and group contact time distributions are also found to be power law and exponential in different time scales. Moreover, the contact arrival process is found to be self similar for data from both our experiment and the Haggle project [4].**

I. Introduction

With the fast technology development on mobile devices, contact based Delay Tolerant Networking (DTN) applications such as mobile peer to peer file sharing which require communication among hand-held devices are gaining more research interests. For example, Nokia N95 smart phones contain an application named *Home Media* where a user is able to share his/her content to peer devices through IEEE 802.11 ad hoc communications. Since one of the main features of DTN is that the network is disconnected most of the time, information is thus exchanged opportunistically among people such that a node only transmits data when it encounters another active device. As a result, understanding human mobility patterns and identifying the social contact dynamics become extremely important. [1]

We have conducted a social contact trace collection experiment similar to [8], [17], [4], [15] [2]. In the experiment, 25 Tmotes [14] with wireless communication capabilities are handed out to an undergraduate class at USC for a week. Each student carries the device during the whole experiment and each mote automatically logs the information about other devices within communication range. Specifically, the time, duration and node IDs of each contact are stored in the memory for off-line processing. In order to maintain high detection probability and reduce the latency of neighbor detection, each node transmits a contact packet every 0.1 second and the nearby nodes receiving the contact packets will record the contact information immediately.

While previous works such as [4], [15] emphasized pairwise dynamics such as contact and inter-contact duration for individual contact, we have not only verified previous observations on our data set but also studied the dynamics of *groups* where a group consists of all the nodes forming a connected networking component at any given time. Intuitively, all the nodes ever involved in a single meeting will be considered as a group. With all the groups identified in the off-line data processing we have investigated the contact and inter-contact time distribution of groups as well as the group properties such as group size and number of group meetings. It is also worth to mention that even though our empirical settings are different from previous ones, the results such as inter-contact time and contact time distribution are quite similar as they fit the same distribution with slightly different parameters.

Self similarity has been identified for ethernet traffic in the 1990s [12], [7]. Recently, researchers have also found out that the contact arrival process in encounter based social networking exhibits self similar nature [5], [18]. In this paper we have examined the data from our experiment as well as the Haggle project and found out that the contact arrival processes in both cases are self similar, with Hurst Parameter (H) equals 0.94 and 0.9 respectively. Self similarity implies the long range dependence and the bursty nature in the contact arrivals, which could potentially provide useful insights on networking protocol designs.

The remainder of the paper is organized as follows. In Section II we present previous relevant works and compare them to our approach. Section III contains the design methodology of the contact collecting experiment, including a sleep cycling mechanism which significantly increases the device lifetime. In Section IV, a figure showing the contact records is given and we present the "contact graph" that illustrates the contact frequency and the grouping effect of the network. The group properties including group sizes, frequency of group meetings are also studied in this section. Then, in Section V, we study

[1] Such studies of human contacts are also useful for building epidemiological models.

[2] The data set from our experiment is available at the following website: http://anrg.usc.edu/www/downloads.

the inter-contact time and contact time distributions, as well as inter-group contact time and group contact time distributions. Section VI contains the study of the self similar nature of contact arrival process for both our experiment and data from Haggle project. Finally, we will give the conclusion and future work directions in Section VII.

II. Related Work

So far, two kinds of experiments have been carried out by researchers in order to study human mobility and contact behaviors. The first approach relies on stable infrastructure access points (AP) such that mobile devices could log the presentation of nearby APs. In the second approach, mobile nodes are only aware of peer devices and encounter based device discoveries are recorded. UCSD [13] and Dartmouth [1] experiments are examples of the first method where participating devices record the visibility of WiFi access points as users move around. The experiments conducted by groups at MIT [8], University of Toronto [17], University of Cambridge [4] and National University of Singapore [15] are good examples of the second approach. In Table I, we summarize all the related experiments mentioned in addition to our study.

As can be seen from Table I, our seven-day experiment focusing on an undergraduate class becomes a complement to the previous contact trace collection experiments. Overall, more than 50000 logs have been recorded by the devices in one week duration. Meanwhile, it is important to notice that as Bluetooth devices cannot provide a sampling interval lower than 10 seconds (since the device discovery phase of Bluetooth communication usually takes more than 10 seconds, and grows linearly as number of neighboring devices increases), we are able to reduce the granularity to 0.1 seconds since Tmotes implement IEEE 802.15.4 radio which could provide almost instant communication. Low sampling rate could ensure a high probability of short term contact detection and maintain low latency on contact discovery, as short term contacts are equally important to long term ones if applications which only require a small amount of data transfer are considered.

Several works [11], [3], [16], [9] have studied the contact and inter-contact time as important factors in mobile ad hoc networks as they determine the delay/throughput of the network, and are critical in designing opportunistic forwarding algorithms. While traditional works modeled the inter-contact time by exponential distribution [11], [16], [19], recent experiments such as [4], [15] show that the distributions of contact and inter-contact time are indeed close to a power law distribution. For example, the authors of [4] examine the data sets from [13], [1] and their own experiment (Haggle) at Cambridge and Intel campus and show that the time between two subsequent contacts follows a power law distribution in a large time scale (about 27 hours). By analyzing the empirical data collected at National University of Singapore, Natarajan *et al.* [15] have looked at more properties including contact time, inter-contact time and inter-pair contact time and all these parameters are found to be power law distributed in certain time scale. In our work, we have not only verified their observations but also studied the grouping effect as well as group dynamics and distributions of inter-group contact time and group contact time.

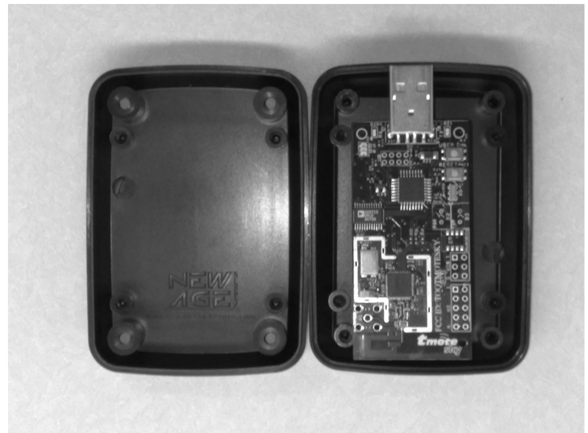

Fig. 1. The mote is placed inside the plastic case which is sealed during the experiment.

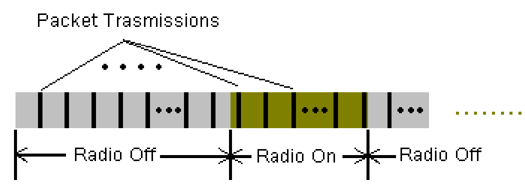

Fig. 2. Figure shows the duty cycling. Note that radio will be explicitly turned on whenever there is a packet transmission.

The self similar nature of contact arrival process in our empirical study as well as Haggle project is also examined by conducting variance-time and R/S (Rescaled Adjusted Range Plot) analysis. In previous works, Chen *et al.* [5] have shown that the visible WiFi access points records in UCSD and Dartmouth experiments [13], [1] are self similar, and Wang *et al.* [18] have also shown that the human contact arrival process is self similar by analyzing traces collected at National University of Singapore.

III. Empirical Methodology

25 volunteers from an undergraduate engineering class have participated in the experiment, each student carries a wireless sensor device (Tmote) with CC2420 radio. To prevent damage, the mote is mounted inside a plastic case, as shown in Figure 1. Devices are handed out in a weekly scheduled class and are collected at the same time in the following week.

Each device periodically broadcasts contact packets every 0.1 second, and every time a node receiving the contact packets will record the time, contact duration and the source node ID. Due to the memory constraint of the wireless device (the flash storage size is about 1MB for Tmote), the contact packet is only going to be logged if two nodes have not seen each other

	Device	Number of devices	Duration	Granularity (seconds)	Number of logs
UCSD	PDA	273	77 days	120	175105
Dartmouth	WiFi Adaptor	approx. 10000	5 years	300	4058284
MIT	Phone	100	9 months	300	N/A
Toronto	PDA	23	16 days	120	2802
Haggle/Intel	iMote	8	3 days	120	3984
Haggle/Infocom	iMote	41	3 days	120	28250
Haggle/Cambridge	iMote	12	5 days	120	8856
Singapore	Phone	12	4 months	30	362599
Our Study	TMote	25	7 days	0.1	53693

TABLE I
SUMMARY AND COMPARISON OF CONTACT TRACE COLLECTING EXPERIMENTS

for more than 10 seconds. Furthermore, at the beginning of the experiment when all the 25 students are sitting together during the class, all devices are set to be inactive to avoid unnecessary memory usage.

Power consumption tests for Tmotes show that the battery can only last for about 4 days if the mote keeps transmitting and receiving packets continuously with 0.1 second interval. In order to enlarge the duration of the experiment to more than 7 days, we have implemented a simple but efficient mechanism where the radio is turned on and off periodically in order to reduce the energy cost, as Figure 2 shows. Specifically, each device will transmit contact packet every 0.1 second with the radio being active for 1.2 second inactive for 1.8 second. Since radio is the main source of energy draw, shutting down the radio in about 60% of the time can significantly reduce the power consumption and the duration of the experiment can be easily extended to more than 7 days.

Among all the 25 devices, 2 of them have been sent back immediately and have not enrolled in the experiment, and 3 motes have been damaged and no records are found in the device memory. Finally, 4 devices contain incomplete logs that have contact information recorded for less than 7 days. Throughout this paper we will focus on the set of nodes that are not damaged and contain contact logs. The recorded contact traces have been extracted and studied after the experiment. The entire contact process during the whole week containing the information of "who meets whom, when, and for how long" can be constructed by aggregating individual logs.

IV. Grouping Effect and Group Properties

A. Grouping Effect

Figure 3 shows the contact logs at a glance. Each horizontal curve represents one node's record with the amplitude of peaks indicating the number of neighboring devices discovered during the experiment. The vertical lines are marks for midnights. The correlated peaks among multiple nodes suggest that event such as class, discussion, etc. are taking place with multiple members involved. Note that Figure 3 only contains

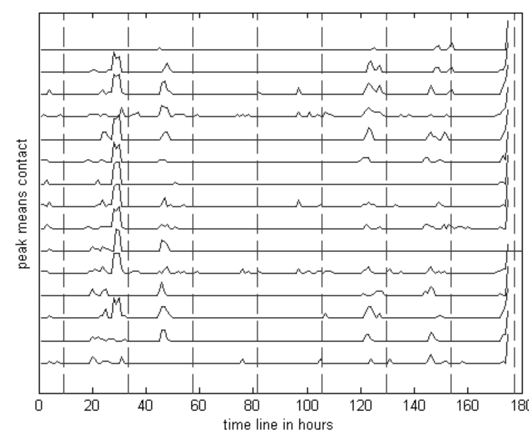

Fig. 3. Experimental data at a glance.

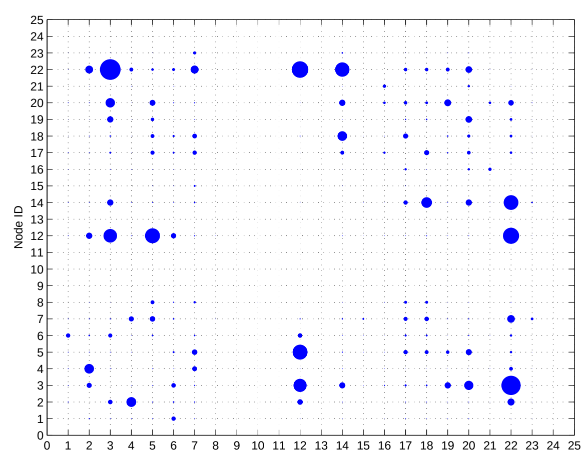

Fig. 4. Strength of pairwise contact. A larger circle size means longer overall contact duration.

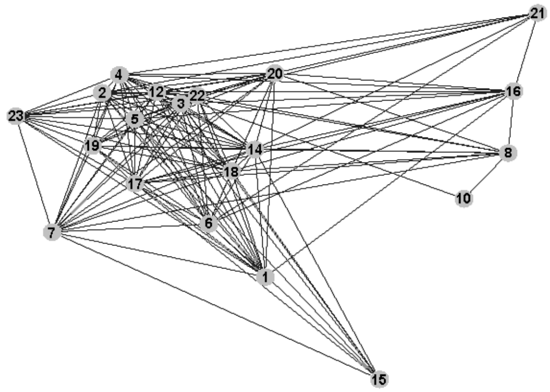

Fig. 5. $G = (V, E)$: A virtual view of the social network (generated using spectral distance embedding).

the devices that have complete 7-day records, and the rest who do not have a complete 7-day log are not presented.

We have studied the pairwise contact strength which measures the overall duration a pair of nodes meet each other during the experiment (Figure 4). It can be seen clearly that some of the pairs such as node 3 and 22 have much longer contact duration compared with others. Some of the nodes have a higher degree due to the fact that they meet more different people during the experiment. These nodes could potentially become good message forwarders due to the fact that they are "connected" to more nodes. Figure 4 also indicates that there exist different group containing different sets of people. For example, node 3, 12 and 22 are strongly connected among each other and thus form a group. Meanwhile, node 12 is closely related to node 5 who is however not related to node 3 and 22, hence node 5 cannot be considered a group member of node 3, 12 and 22. In this case, node 12 and node 5 should be considered as a different group.

In order to better visualize the grouping effect, a graph $G = (V, E)$ is generated where V denotes the set of all nodes that have contact information recorded during the experiment (i,e.: node 9, 11 and 13 with no available logs are ignored) , and E denotes the set of edges connecting the vertices (as shown in Figure 5). An edge $e = (u, v)$ indicates that node u and v have ever been in contact during the experiment, with the distance of e indicating how "close" two nodes are related, i.e.: the closer the two vertices are, the longer they have stayed together during the experiment. The graph is drawn with the Spectral Distance Embedding (SDE) technique [6], which is essentially a variation of Multidimensional Scaling mechanism [2]. By comparing Figure 5 with Figure 4, it is clear that the nodes close to the perimeter are the ones that appear less frequently in communication, as compared to the central nodes who have made extensive contact with others.

To be more precise, a *group* is defined by the set of nodes

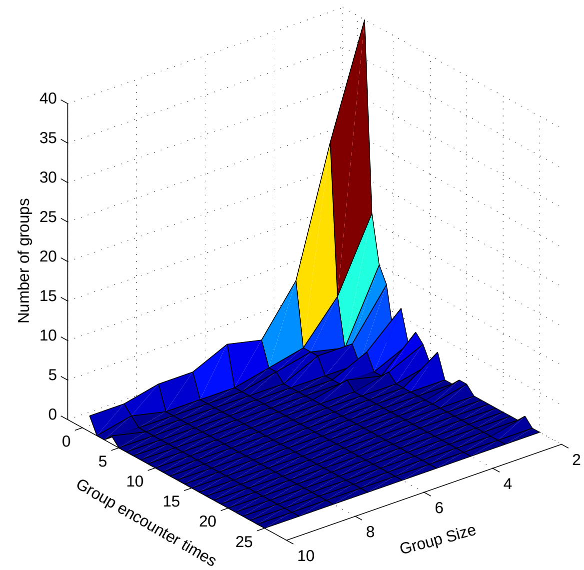

Fig. 6. Number of groups satisfying certain properties (group size and number of meetings)

Group size	2-3	4-6	7-10	> 10
Number of such groups	137	28	9	9

TABLE II
DISTRIBUTION OF GROUP SIZE.

that form a connected network component at any given time during the experiment. It has been observed that many different groups may possibly exist during the experiment, and in the next subsection we will study group properties including the distributions of group size and number of group meetings.

B. Group Properties

Identifying the grouping effect as well as the group properties could provide useful insights for protocol design in interaction based mobile networks. For example, a routing protocol may need to identify a "group leader" and rely on that node to forward messages from the outside network to members within the group. Thus it is crucial to understand metrics such as number of groups, number of nodes in a group and the frequency that a group of people gather together.

177 distinct groups have been discovered in our contact study experiment with group size varying from 2 to 14 and number of group meetings varying from 1 to 43. Figure 6 plots the number of groups with respect to group size and number of group encounters. It can be seen that there exists a large amount of "instant groups" that have shown up only

Group meeting times	1	2-5	6-10	> 10
Number of such groups	100	57	20	7

TABLE III
DISTRIBUTION OF GROUP MEETING TIMES.

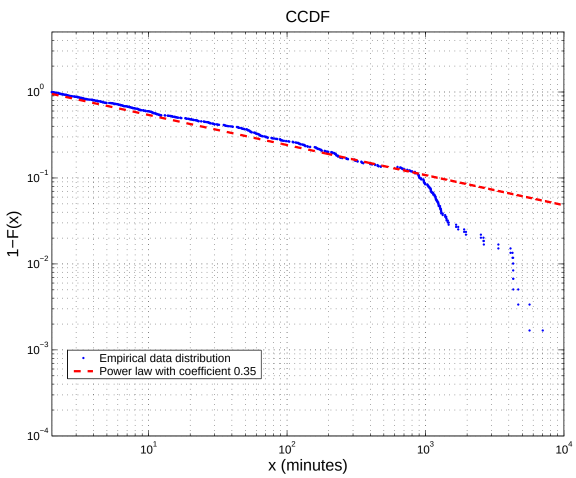

Fig. 7. Inter-contact time distribution (log scale).

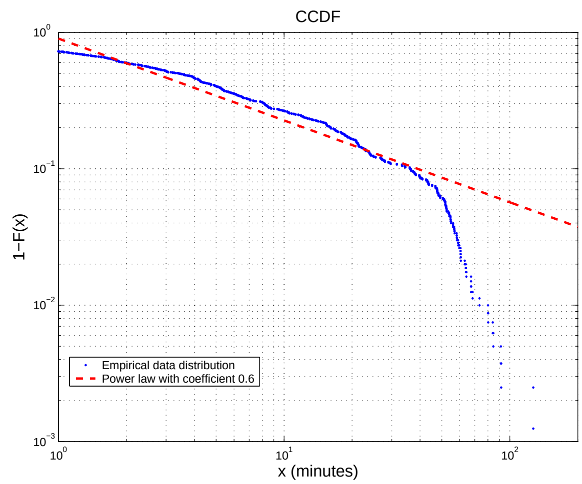

Fig. 8. Pairwise contact time distribution (log scale).

once during the whole experiment. Although most of groups have small sizes and meeting times, there exists larger ones who meanwhile have met more than once. For example, a group of 10 people have met four times during the experiment (as can be found in Figure 6). The statistics of number of groups satisfying different criteria are also given in Table II and Table III, which summarize the results in Figure 6.

V. Important Metrics and Modeling

A. Inter-contact Time and Contact Time Distribution

Researchers have studied the inter-contact and contact duration for interaction based social networks in previous works, and both are found to be power law distributed over certain range. Specifically, in Haggle IMote experiment [4], it is observed that the power law coefficient $k = 0.5$ for inter-contact duration and $k = 1.5$ for contact duration. In NUS data [15], the inter-contact duration exhibits a power law with $k = 0.55$ and $k = 0.84$ for contact duration. We have examined the same metrics and the results can be found in Figure 7 and Figure 8. Similarly, both distributions appear to be power law in certain time scale, and the power law coefficient k is equal to 0.35 for inter-contact duration and 0.6 for contact duration. The contact duration distribution decays slower as compared to previous studies, and the reason is that in general, students tend to stay together longer as they may have same classes/discussions. Moreover, the inter-contact time distribution also decays slowly with coefficient $k = 0.35$. Besides the fact that our experiment focuses on less number of people which leads to a lower chance of node encounter, a possible reason could be the infrequent interactions among these students after class.

B. Inter-group Contact Time and Group Contact Time Distribution and Model

In addition, we have investigated the distribution of inter-group contact time as well as group contact time. Inter-group contact time is defined by the time between subsequent group meetings, which infers the time that a message has to stay inside a group in order to be sent out. Group contact time is the duration that a group stays together, which determines how much data group members are able to exchange. Understanding the distributions of these metrics will also be very helpful in protocol design aspect. An example is that a group header node implementing sleep scheduling mechanism has to determine how often to wake up in order to maintain high probability of node discovery.

The complementary cumulative density function (ccdf) of inter-group contact time as well as group contact time distributions are plotted in Figure 9 and Figure 10. Similarly, the distributions are found to be pareto (heavy-tailed) in certain time scale and then decays exponentially. Let $f(t)$ be the probability density function of inter-group contact time distribution, then $f(t)$ can be approximated as

$$f(t) = \begin{cases} k\frac{\tau^k}{t^{k+1}}, & \text{if } t < t_c \\ \lambda e^{-\lambda t}, & \text{if } t > t_c \end{cases}$$

That is, the distribution follows a power law and an exponential distribution in different time scales cut off by t_c. Note that $\tau = 1min$ is arbitrarily chosen as the time unit that determines the minimum possible inter-contact time. By curve fitting, it can be found out that $t_c = 2000$, $k = 0.23$ and $\lambda = 0.0006$. The reason that a smaller k value is observed compared to inter-contact time distribution is that it takes longer for a group to "re-union" compared to the case where the time between any two contacts is taken into consideration, thus on average the time between two group meetings is longer.

Similarly, the probability density function $g(t)$ for group contact time distribution can be approximated as

$$g(t) = \begin{cases} k'\frac{\tau^{k'}}{t^{k'+1}}, & \text{if } t < t'_c \\ \lambda' e^{-\lambda' t}, & \text{if } t > t'_c \end{cases}$$

where $t'_c = 60$, $k' = 0.8$ and $\lambda' = 0.06$ based on best curve fittings. As compared to the inter-group contact time, the group

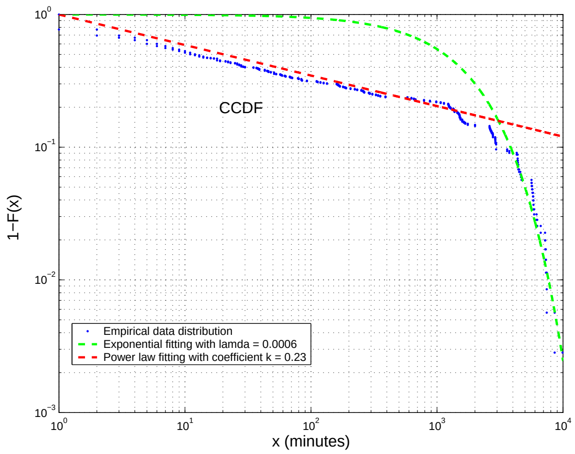

Fig. 9. Inter-group contact time distribution (log scale).

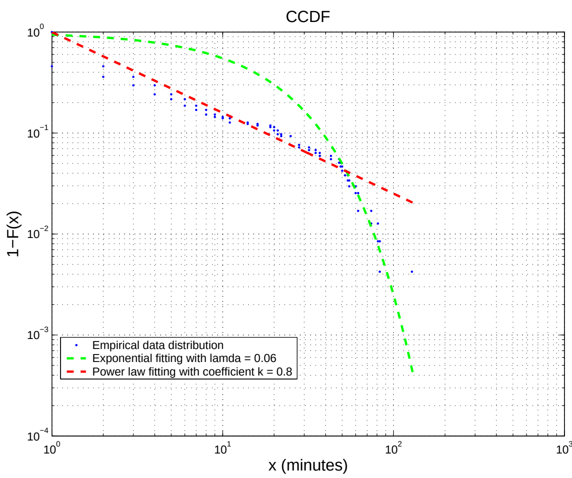

Fig. 10. Group contact time distribution (log scale).

contact time distribution follows the power law in a much smaller time scale (60 minutes) and decays much faster with the slope k' close to 0.8.

VI. Self Similarity

We examine the data collected in Haggle project [4] [3] as well as our empirical study and apply variance-time and R/S plots (Rescaled Adjusted Range Plot) in order to verify the self similarity of the contact arrival process. It turns out that the two empirical studies have very close behavior in terms of self similarity, with Hurst Parameter $H \approx 0.9$ in Haggle project and $H \approx 0.94$ in our experiment. The self similarity of contact arrival process not only indicates the long range dependence of human mobility and contact behaviors, but also suggests the bursty nature of the contact arrivals.

Figure 11 and Figure 13 show that the variance of the aggregated contact process decays slowly as the time scale

[3] We focus on the Haggle experiment conducted during the IEEE INFOCOM 2005 conference. Data is available at [1].

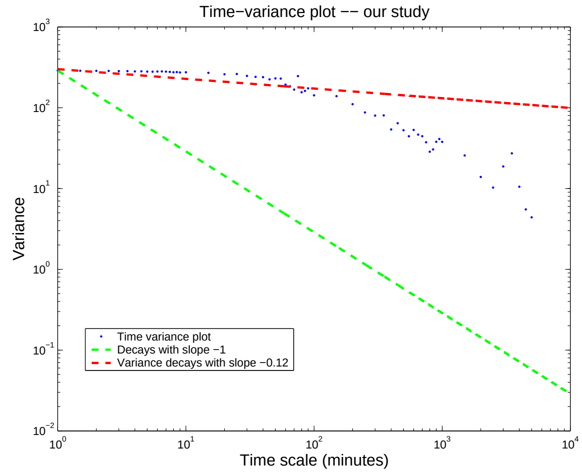

Fig. 11. Figure shows that variance decays slowly as m increases in our experiment.

increases. Note that the decay of variance can be expressed by $Var(X^{(m)}) = a \cdot m^{-\beta}$ in which a is constant, m is the time scale length and β is a value ranging from 0 to 1 which describes how fast the variance decays. For general processes such as Poisson Process, $\beta = 1$. For self similar processes, $\beta < 1$ and the smaller β is, the larger effect of long range dependence the process has and hence is more "self similar". Since the relation between Hurst Parameter and variance decay slope satisfies $H = 1 - \frac{\beta}{2}$ [12], we are able to conclude that $H \approx 0.9$ and $H \approx 0.94$ for Haggle project and our experiment respectively.

The R/S analysis is another useful tool to identify self similarity which divides the process by logarithmical intervals and calculates the rescaled adjusted range for each subinterval, and finally takes the average of all calculated values [10]. H can be estimated by the slope of R/S plot. Figure 12 and Figure 14 give an estimation of $H \approx 0.9$ for Haggle experiment and $H \approx 0.94$ for our study, which verify our conclusion above.

VII. Conclusions and Future Work

We have conducted a human contact trace collection experiment which serves as a complement to previous relevant studies. We are able to reduce the probing frequency to 0.1 second which significantly increases the chance of detection of short contact and decreased the latency of neighbor discovery. Data containing contact information is processed off-line. We have identified the grouping effect and studied group properties including the distribution of group size and frequency of group meetings. Inter-contact time and contact time, as well as inter-group contact time and group contact time distributions have been studied and modeled. Surprisingly, our observations are similar to previous findings even though the empirical settings are different. For example, the inter-contact time and contact time fit the same distribution with different parameters. This could help researchers develop suitable simulation models for

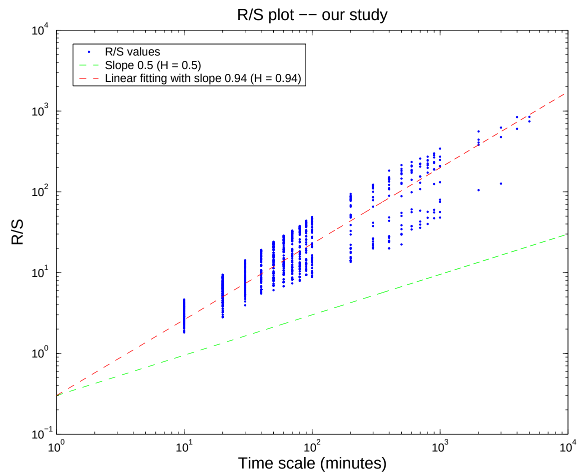

Fig. 12. R/S plot for our experiment.

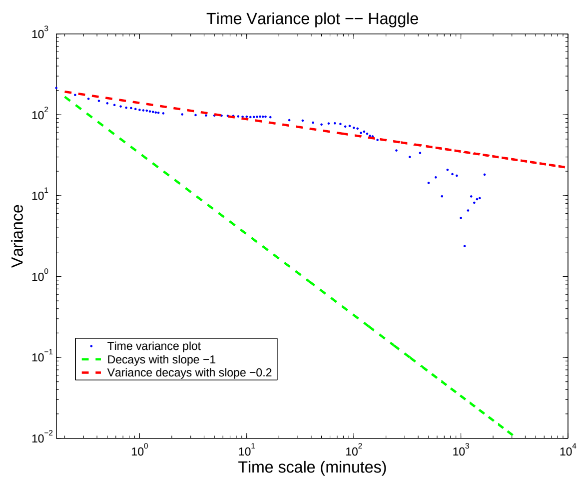

Fig. 13. Figure shows that variance decays slowly as m increases in Haggle.

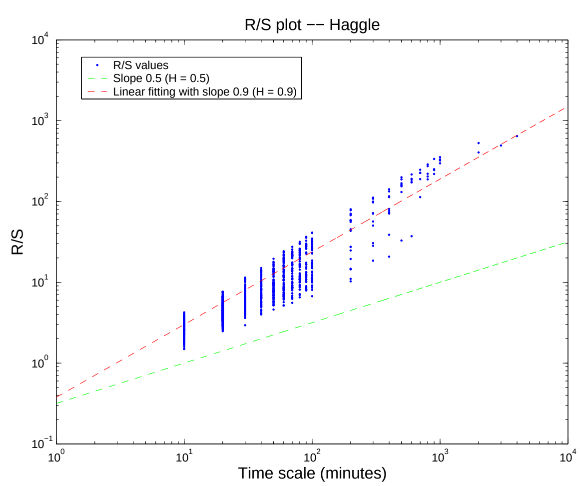

Fig. 14. R/S plot for Haggle.

human contact based networks. Finally we have investigated the self similarity of contact arrival process by conducting variance-time and R/S analysis to the data from Haggle project and our own experiment.

For future works, we plan to carry out a larger scale contact study experiment with longer duration such that the dynamics of human interaction based networks can be more extensively examined. We also plan to design and implement energy efficient algorithms on mobile devices based on our empirical findings and investigate the tradeoffs afterwards.

REFERENCES

[1] Crawdad project, http://crawdad.cs.dartmouth.edu.

[2] H. Abdi. Metric multidimensional scaling (mds): Analyzing distance matrices. In *N.J. Salkind (Ed.): Encyclopedia of Measurement and Statistics*, Thousand Oaks (CA): Sage.

[3] A. Al-Hanbali, A. A. Kherani, R. Groenevelt, P. Nain, and E. Altman. Impact of mobility on the performance of relaying in ad hoc networks. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–12, 2006.

[4] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding. Technical Report 617, University of Cambridge, February 2005.

[5] L. Chen, Y. Chen, T. Sun, P. Sreedevi, K. Chen, C. Yu, and H. Chu. Finding self-similarities in opportunistic people networks. In *In IEEE Infocom*, 2007.

[6] A. Civril, M. Magdon-ismail, and E. Bocek-rivele. Sde: Graph drawing using spectral distance embedding. In *In Proc. Graph Drawing*, pages 512–513, 2005.

[7] M. E. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE /ACM Transactions on Networking*, 5(6):835–846, 1997.

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

[9] Sharma. G, Mazumda. R, and Shroff. N. B. Delay and capacity trade-offs in mobile ad hoc networks: A global perspective. In *Proceedings of IEEE INFOCOM*, Barcelona, Catalunya, Spain, August 2006.

[10] M. Gospodinov and E. Gospodinova. The graphical methods for estimating hurst parameter of self-similar network traffic. In *In ICCST*, 2005.

[11] M. Grossglauser and D. N. C. Tse. Mobility increases the capacity of ad-hoc wireless networks. In *INFOCOM*, pages 1360–1369, 2001.

[12] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic (extended version). *Networking, IEEE/ACM Transactions on*, 2(1):1–15, 1994.

[13] M. Mcnett and G. M. Voelker. Access and mobility of wireless pda users. *SIGMOBILE Mob. Comput. Commun. Rev.*, 9(2):40–55, April 2005.

[14] moteiv. Telosb motes datasheet. http://www.moteiv.com/products/docs/telos-revb-datasheet.pdf.

[15] A. Natarajan, M. Motani, and V. Srinivasan. Understanding urban interactions from bluetooth phone contact traces. In *PAM 2007, 8th Passive and Active Measurement conference*, pages 115–124, Louvain-la-neuve, Belgium, April 2007.

[16] G. Sharma and R. R. Mazumdar. In *Workshop on Modeling and Optimization in Mobile Ad Hoc and Wireless Networks*, Cambridge, UK, March 2004.

[17] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara. User mobility for opportunistic ad-hoc networking. In *Mobile Computing Systems and Applications, 2004. WMCSA 2004. Sixth IEEE Workshop on*, pages 41–50, 2004.

[18] W. Wang, V. Srinivasan, and M. Motani. Adaptive contact probing mechanisms for delay tolerant applications. In *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 230–241, Montreal, Quebec, Canada, 2007. ACM Press.

[19] Y. Wang, S. Kapadia, and B. Krishnamachari. Infection spread in wireless networks with random and adversarial node mobilities. In *Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models*, HongKong, China, May 2008.