Eagle, N., and Pentland, A., "Reality Mining: Sensing Complex Social Systems"',
J. of Personal and Ubiquitous Computing. To appear: June 2005.

Reality Mining: Sensing Complex Social Systems

Nathan Eagle and Alex (Sandy) Pentland

MIT Media Laboratory
20 Ames St.
Cambridge, MA 02139
{nathan, sandy}@media.mit.edu

Abstract. We introduce a system for sensing complex social systems with data collected from one hundred mobile phones over the course of six months. We demonstrate the ability to use standard Bluetooth-enabled mobile telephones to measure information access and use in different contexts, recognize social patterns in daily user activity, infer relationships, identify socially significant locations, and model organizational rhythms.

1 Introduction

The last ten years could rightly be coined the decade of the mobile phone. In 2004, over 600 million handsets were sold, dwarfing the number of personal computers sold that year [22]. The potential functionality of this ubiquitous infrastructure of mobile devices is dramatically increasing. In this paper we describe how data collected from mobile phones can be used to uncover regular rules and structure in behavior of both individuals and organizations. We begin with a discussion of the rationale for using phones as wearable sensors and the type of data they can collect. Subsequently we describe the benefits of fusing information from cell towers with discovered Bluetooth IDs, and incorporate this into models of individual users. The initial results of our ongoing user study on phone usage and communication patterns are discussed. Turning our attention away from individuals and toward dyads, we extract salient features indicative of the relationships between subjects using proximity, time, and location data. Finally, with the nodes and edges of this social network identified, the concept of organizational rhythms is introduced as useful metric for quantifying organizational behavior.

2 Mobile Phones as Wearable Sensors

For over a century social scientists have conducted surveys to learn about human behavior. Surveys are plagued with issues however, such as bias, sparsity of data, and lack of continuity between discrete questionnaires. It is this absence of dense, continuous data that also hinders the machine learning and agent-based modeling communities from constructing more comprehensive predictive models of human dynamics. Over the last two decades these has been a significant amount of research attempting to address these issues by building location-aware devices capable of collecting rich behavioral data [18]. While these projects were relatively successfully, by depending on a limited supply of custom hardware, they were unable to scale to groups of greater size. However, with the rapid technology adoption of mobile phones comes an opportunity to collect a much larger dataset on human behavior [7, 13]. The very nature of mobile phones makes them an ideal vehicle to study both individuals and organizations: people habitually carry their mobile phones with them and use them as a medium for much of their communication. In this paper we capture all the information to which the phone has access (with the exception of content from phone calls or text messages) and describe how it can be used to provide insight into both the individual and the collective.

2.1 Mobile Phone Proximity Logs

One of the key ideas in this paper is to exploit the fact that modern phones use both a short-range RF network (e.g., Bluetooth) and a long-range RF network (e.g., GSM), and that the two networks can augment each other for location and activity inference. The idea of logging cell tower ID to determine approximate location will be familiar to readers, but the idea of logging Bluetooth devices is relatively recent and provides very different types of information [11].

Bluetooth is a wireless protocol in the 2.40-2.48 GHz range, developed by Ericsson in 1994 and released in 1998 as a serial-cable replacement to connect different devices. Although market adoption has been initially slow, according to industry research estimates, by 2006 90% of PDAs, 80% of laptops, and 75% of mobile phones will be shipped with Bluetooth [23]. Every Bluetooth device is capable of 'device-discovery', which allows them to collect information on other Bluetooth devices within 5-10 meters. This information includes the Bluetooth MAC address (BTID), device name, and device type. The BTID is a hex number unique to the particular device. The device name can be set at the user's discretion; e.g., "Tony's Nokia". Finally, the device type is a set of three integers that correspond to the device discovered; e.g., Nokia mobile phone, or IBM laptop.

To log BTIDs we designed a software application, BlueAware, that runs passively in the background on MIDP2-enabled mobile phones. Bluetooth was primarily designed to enable wireless headsets or laptops to connect to phones, but as a by-product, devices are becoming aware of other Bluetooth devices carried by people nearby. Our application records and timestamps the BTIDs encountered in a proximity log and makes them available to other applications, similar to the Jabberwocky project developed by Paulos et al. [14]. BlueAware is automatically run in the background when the phone is turned on, making it essentially invisible to the user.

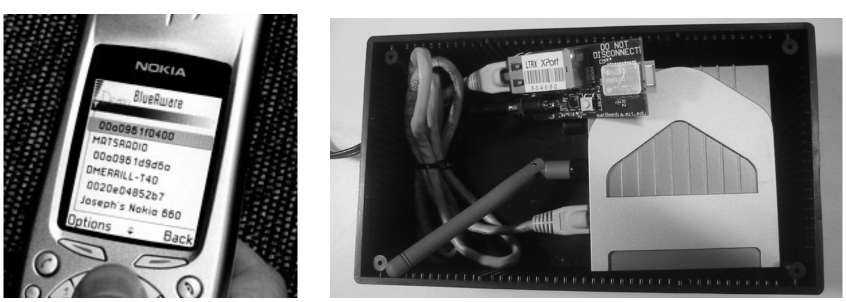

Fig. 1 Methods of detecting Bluetooth devices. BlueAware running in the foreground on a Nokia Series 60 phone (left). Bluedar, a Bluetooth beacon coupled with a WiFi bridge (right).

A variation on BlueAware is Bluedar. Bluedar was developed to be placed in a social setting and continuously scan for visible devices, wirelessly transmitting detected BTIDs to a server over an 802.11b network. The heart of the device is a Bluetooth beacon designed by Mat Laibowitz incorporating a class 2 Bluetooth chipset that can be controlled by an XPort web server [10]. We integrated this beacon with an 802.11b wireless bridge and packaged them in an unobtrusive box. An application was written to continuously telnet into multiple BlueDar systems, repeatedly scan for Bluetooth devices, and transmit the discovered proximate BTIDs to our server. Because the Bluetooth chipset is a class 2 device, it is able to detect any visible Bluetooth device within a working range of up to twenty-five meters. We are currently using the system to prototype a proximity-based introduction service [6].

Refresh Rate vs. Battery-Life. Continually scanning and logging BTIDs can expend an older mobile phone battery in about 18 hours.[1] While continuous scans provide a rich depiction of a user's dynamic environment, most individuals expect phones to have standby times exceeding 48 hours. Therefore BlueAware was modified to only scan the environment once every five minutes, providing at least 36 hours of standby time.

2.2 Privacy Implications

Mining the reality of our one hundred users raises justifiable concerns over privacy. However, the work in this paper is a social science experiment, conducted with human subject approval and consent of the users. Outside the lab we envision a future where phones will have greater computation power and will be able to make relevant inferences using only data available to the user's phone. In this future scenario, the inferences are done in real-time on the local device, making it unnecessary for private information to be taken off the handset. However, the computational models we are currently using cannot be implemented on today's phones. Thus, our results aim to show the potential of the information that can be gleaned from the phone, rather than presenting a system that can be deployed today outside the realm of research.

2.3 The Dataset

Our study consists of one hundred Nokia 6600 smart phones pre-installed with several pieces of software we have developed as well as a version of the Context application from the University of Helsinki [15]. Seventy-five users are either students or faculty in the MIT Media Laboratory, while the remaining twenty-five are incoming students at the MIT Sloan business school adjacent to the laboratory. Of the seventy-five users at the lab, twenty are incoming masters students and five are incoming MIT freshman. The information we are collecting includes call logs, Bluetooth devices in proximity, cell tower IDs, application usage, and phone status (such as charging and idle), which comes primarily from the Context application. The study will generate data collected by one hundred human subjects over the course of nine months and represent approximately 500,000 hours of data on users' location, communication and device usage behavior.[2] Upon completion of the study, we plan to release a public, anonymous version of the data set for other researchers to use.

---

[1] Using a 6-month old battery of a Nokia 6600 in a sparsely populated Bluetooth environment

[2] At the time of submission one hundred human subjects have been participating in the study for time periods ranging from two to seven months, representing over 250,000 hours of data.

3 User Modeling: Identifying Structure in Routine

Although humans have the potential for relatively random patterns of behavior, there are easily identifiable routines in every person's life. These can be found on a range of timescales: from the daily routines of getting out of bed, eating lunch, and driving home from work, to weekly patterns such as the Saturday afternoon softball games, to yearly patterns like seeing family during the holidays in December. While our ultimate goal is to create a predictive classifier that can perceive aspects of a user's life more accurately than a human observer (including the actual user), we begin by building simple mechanisms that can recognize many of the common structures in the user's routine. Learning the structure of an individual's routine has already been demonstrated using other modalities, however we present this analysis as a foundation which will then be extended to demonstrate the learning of *social* structures.

We begin with a simple model of behavior in three states: home, work, and elsewhere. The data are obtained from Bluetooth, cell tower, and temporal information collected from the phones. We then incorporate information from static Bluetooth devices (class 1, such as desktop computers), using them as 'cell towers' to identify significant locations and localize the user to a ten meter radius. We show that most users spend a significant amount of time in the presence of static Bluetooth devices, particularly when they don't have cell tower reception (e.g., inside the office building). This makes them an ideal supplement to cell towers for location classification.

3.1 Location based on cell towers and Bluetooth

There has been a significant amount of research which correlates cell tower ID with a user's location [2, 3, 8]. For example, Laasonen et al. describe a method of inferring significant locations from cell tower information through analysis of the adjacency matrix formed by proximate towers. They were able to show reasonable route recognition rates, and most importantly, succeeded in running their algorithms directly on the mobile phone [9].

Obtaining accurate location information from cell towers is complicated by the fact that phones can detect cell towers that are several miles away. Furthermore, in urban areas it is not uncommon to be within range of more than a dozen different towers. The inclusion of information about all the current visible towers as well as their respective signal strengths would help solve the location classification problem, although multipath distortion may still confound estimates.

We observe that relatively high location accuracy may also be achieved if the user spends enough time in one place to provide an estimate of the cell tower *probability density function*. Phones in the same location can be connected to different cell towers at different times depending on a variety of variables including signal strength and network traffic. Thus, over time each phone 'sees' a number of different cell towers, and the distribution of detected towers can vary substantially with even small changes in location. Figure 2 shows the distribution of cell towers seen for a given area with a 10m radius. Towers were only included in these distributions if the common area's static Bluetooth desktop computer was also visible, ensuring the users' location within 10m (or less). Discrepancies in the distributions are attributed to the users' typical position within the 10m radius. Users 2 and 4 both share a window office and have virtually the same cell tower distribution, despite having a very different distribution of hours spent in the office (as verified by the Bluetooth and cell tower logs). Users 1

The total duration of the study will be for nine months, and all users will have been enrolled for at least six months.

and 5 both spend the majority of their time in the common area away from the windows and see only half as many towers as the others. User 3 is in a second office in the same area, and has a distribution of cell towers that is intermediate between the two other sets of users.

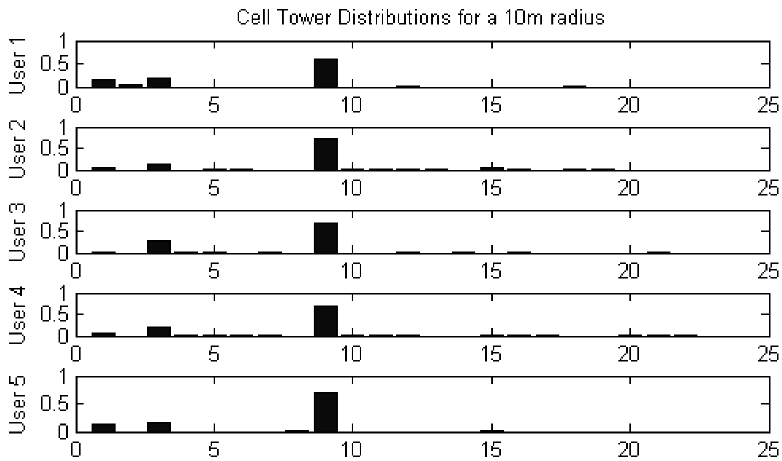

Fig 2. The probability distribution of seeing twenty-five cell towers from the third floor corner of an office building using 150 hours of data from each of five users. (Ranged was assured to 10m by the presence of a static Bluetooth device.)

Despite progress in mapping cell tower to location, the resolution simply cannot be as high as many location-based services require. GPS is an alternative approach that has been used for location detection and classification [1, 12, 19], but the line-of-sight requirements prohibit it from working indoors. We have therefore incorporated the use of static Bluetooth device ID as an additional indicator of location, and shown that it provides a significant improvement in user localization, especially within office environments. This fusion of data is particularly appropriate since areas where cellular signals are weak, such as in the middle of large buildings, often correspond to places where there are many static Bluetooth devices, such as desktop computers. On average, the subjects in our study were without mobile phone reception 6% of the time. When they did not have reception, however, they were within range of a static Bluetooth device or another mobile phone 21% and 29% of their time, respectively. We expect coverage by Bluetooth devices to increase dramatically in the near future as they become more common in computers and electronic equipment.

We believe Bluetooth ID may become as important as cell tower mapping for estimation of user location. Figure 3 below shows the ten most frequently detected Bluetooth devices for one subject averaged over the month of January. This figure not only provides insight into the times the user is in his office (from the frequencies of the top 'Desktop'), but as mentioned in Section 4, also the type of relationship with other subjects. For example, the figure suggests the user leaves his office during the hour of 14:00 and becomes increasingly proximate to Subject 4. Judging from the strong cutoffs at 9:00 and 17:00, it is clear this subject had very regular hours during the month, and thus has fairly predictable high-level behavior. This "low entropy" behavior is also depicted in Figure 4.

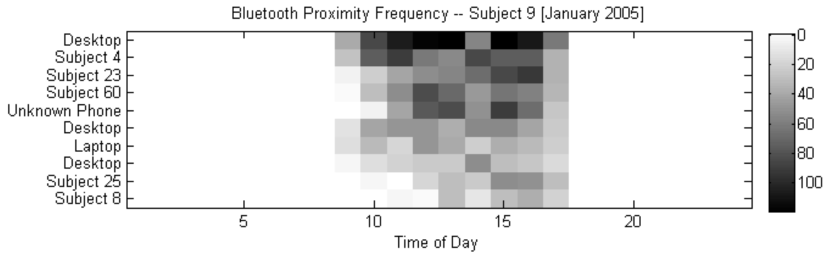

Fig 3. The number of Bluetooth encounters for Subject 9 over the month of January

3.2 Models to Identify Location and Activity

Human life is inherently imbued with routine across all temporal scales, from minute-to-minute actions to monthly or yearly patterns. Many of these patterns in behavior are easy to recognize, however some are more subtle. We attempt to quantify the amount of predictable structure in an individual's life using an entropy metric. People who live high-entropy lives tend to be more variable and harder to predict, while low-entropy lives are characterized by strong patterns across all time scales. Figure 4 depicts the patterns in cell tower transitions and the total number of Bluetooth devices encountered each hour during the month of January for Subject 9, a 'low entropy' subject.

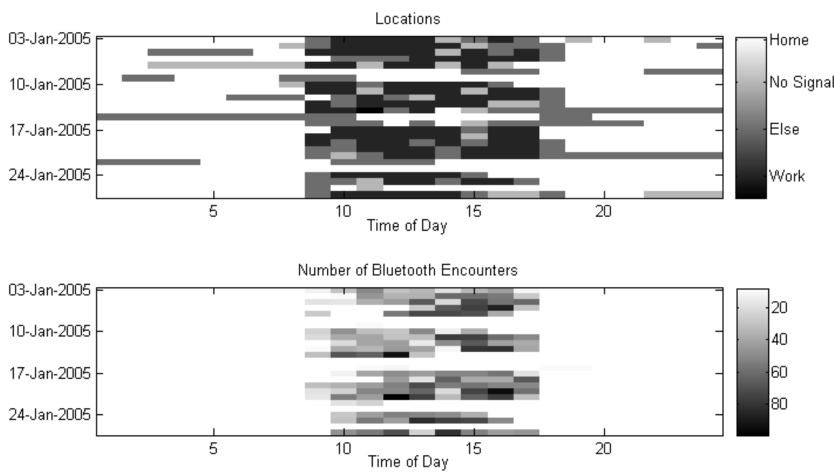

Fig 4. Subject 9's 'low entropy' daily distribution of home/work transitions and Bluetooth devices. The 'hot spot' in mid-day is when the subject is at the workplace.

It is clear that the subject is typically at home during the evening and night until 8:00, when he commutes in to work, and then stays at work until 17:00 when he returns home. We can see that almost all of the Bluetooth devices are detected during these regular office hours, Monday through Friday. This is certainly not the case for many of the subjects. Figure 5 displays a different set of behaviors for Subject 8. The subject has much less regular patterns of location and in the evenings has other mobile devices in close proximity. We will use contextualized information about proximity with other mobile devices to infer relationships, described in section 4.2.

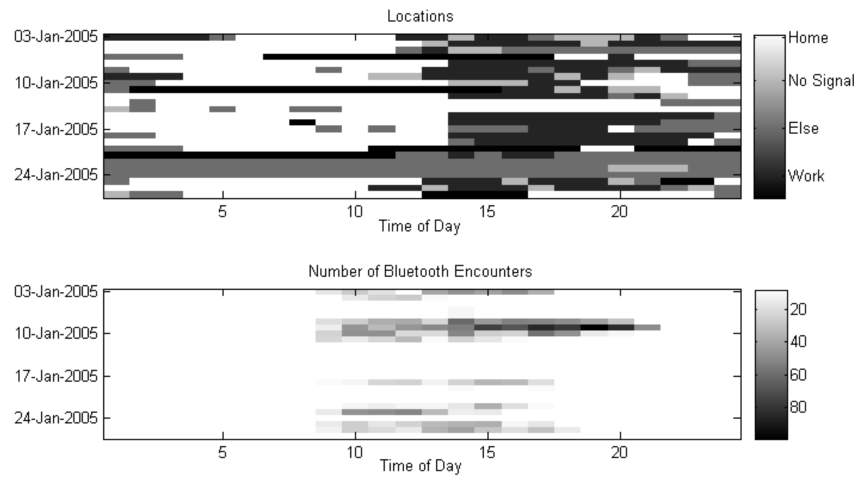

Fig 5. Subject 4's 'high entropy' daily distribution of home/work transitions and Bluetooth devices.

One similarity between the two different behaviors above is the clear role time plays in determining user behavior. To account for this, we have developed a simple Hidden Markov Model conditioned on both the hour of day as well as weekday or weekend. A straightforward Expectation-Maximization inference engine was used to learn the parameters in the model, and performed clustering in which we defined the dimensionality of the state space. After training our model with one month of data from several subjects we were able to provide a good separation of ({*office*}, {*home*}, {*elsewhere*}) clusters, typically with greater than 95% accuracy. Examination of the data shows that non-linear techniques will be required to obtain significantly higher accuracy. However, for the purposes of the next two sections, this accuracy has proven sufficient. In future work we hope to leverage the information within LifeNet [17] to create more specific interferences about activity.

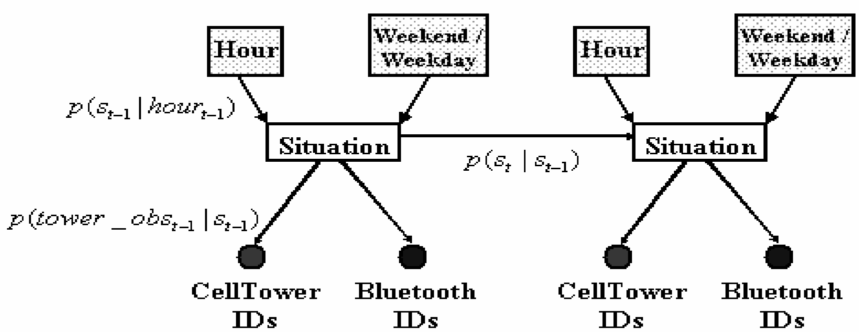

Fig 6. A Conditioned Hidden Markov Model for situation identification. The model was designed to be able to incorporate many additional observation vectors such as devices nearby, traveling, sleeping and talking on the phone.

3.4 Mobile Usage Patterns in Context

Capturing mobile phone usage patterns of one hundred people for an extended period of time can provide insight into both the users as well and the ease of use of the device itself. For example, 35% of our subjects use the clock application on a regular basis (primarily to set the alarm clock and then subsequently to press snooze) yet it takes 10 keystrokes to open the application from the phone's default settings. Not surprisingly, specific applications, such as the alarm clock, seem to be used much more often at home rather than at work. Figure 7 is a graph of the aggregate popularity of the following applications when both at home and at work. It is interesting to note that despite the subjects being technically savvy, there was not a significant amount of usage of the sophisticated features of the phone - indeed the default game "Snake" was used just as much as the elaborate Media Player application.

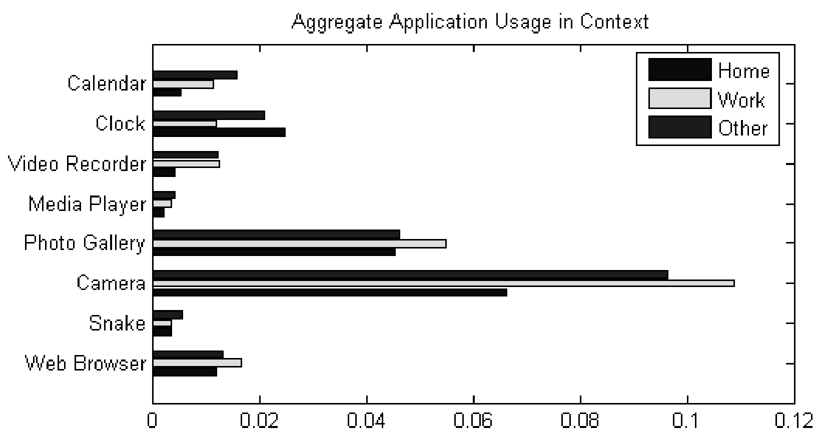

Fig 7. Average application usage of 100 subjects with location

While there is much to be gained from a contextual analysis of new application usage, perhaps the most important and still most popular use of the mobile phone is as a communication device. Figure 8 is a break down of the different types of usage patterns from a selection of the subjects. Approximately 81% of communication on the phone was completed by placing or receiving a voice call. Data (primarily email) was at 13% of the communication, while text messaging was 5%.

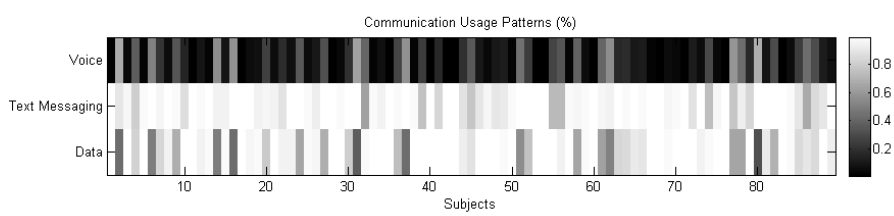

Fig 8. Average communication mediums for a sample of 90 subjects

Learning user's application routines can enable the phone to place a well-used application in more prominent places, for example, as well as creating a better model of the behavior of an individual [20]. As we shall show in Section 4, these models can also be augmented with additional information about a user's social context.

3.5 Data Characterization and Validation

The following section describes how errors may be introduced into the data through data corruption, device detection failures, and most significantly, through human error.

Data Corruption. All the data from a phone are stored on a flash memory card, which has a finite number of read-write cycles. Initial versions of our application wrote over the same cells of the memory card. This led to failure of a new card after about a month of data collection, resulting in the complete loss of data. When the application was changed to store the incremental logs in RAM and subsequently write each complete log to the flash memory, our data corruption issues virtually vanished. However, ten cards were lost before this problem was identified, destroying portions of the data collected during the months of September and October for six Sloan students and four Media Lab students.

Bluetooth Errors. One central intent of this research is to verify the accuracy of automatically collected data from mobile phones for quantifying social networks. We are facing several technical issues. The ten meter range of Bluetooth along with the fact that it can penetrate some types of walls, means that people not physically proximate may incorrectly be logged as such. By scanning only periodically every five minutes, shorter proximity events may also be missed.

Additionally, there is a small probability (between 1-3% depending on the phone) that a proximate, visible device will not be discovered during a scan. Typically this is due to either a low level Symbian crash of an application called the "BTServer", or a lapse in the device discovery protocol. The BT server crashes and restarts approximately once every three days (at a 5 minute scanning interval) and accounts for a small fraction of the total error. However, to detect other subjects, we can leverage the redundancy implicit in the system. Because both of the subjects' phones are actually scanning, the probability of a simultaneous crash or device discovery error is less than 1 in 1000 scans.

In our tests at MIT, we have empirically found that these errors have little effect on the extremely strong correlations between interaction (survey data) and the 10m Bluetooth proximity information. These problems therefore produce a small amount of 'background noise' against which the true proximity relationships can be reasonably measured. However, social interactions within an academic institution are not necessarily typical of a broader cross-section of society and the errors may be more severe or more patterned. If testing in a more general population shows that the level of background noise is unacceptable, there are various technical remedies available. For instance, the temporal pattern of BTID logs allows us to identify various anomalous situations. If someone is not involved in a specific group conversation but just walking by, then they will often enter and leave the log at a different time than the members of the group. Similar geometric and temporal constraints can be used to identify other anomalous logs.

Human Induced Errors. The two primary types of human-induced errors in this dataset result from the phone either being off, or separated from the user. The first error comes from the phone being either explicitly turned off by the user or exhausting the batteries. According to our collected survey data, users report exhausting the batteries approximately 2.5 times each month. One fifth of our subjects manually turn the phone off on a regular basis during specific contexts such as classes, movies, and (most frequently) when sleeping. Immediately before the phone powers down, the event is timestamped and the most recent log is closed. A new log is created when the phone is restarted and again a timestamp is associated with the event.

A more critical source of error occurs when the phone is left on, but not carried by the user. From surveys, we have found that 30% of our subjects claim to never forget their phones, while 40% report forgetting it about once each month, and the remaining 30% state that they forget the phone approximately once each week. Identifying the times where the phone is on, but left at home or in the office presents a significant challenge when working with the dataset. To grapple with the problem, we have created a 'forgotten phone' classifier. Features included staying in the same location for an extended period of time, charging, and remaining idle through missed phone calls, text messages and alarms. When applied to a subsection of the dataset which had corresponding diary text labels, the classifier was able to identify the day where the phone was forgotten, but also mislabeled a day when the user stayed home sick. By ignoring both days, we risk throwing out data on outlying days, but have greater certainty that the phone is actually with the user. A significantly harder problem is to determine whether the user has temporarily moved beyond ten meters of his or her office without taking the phone. Empirically, this appears to happen with many subjects on a regular basis and there doesn't seem to be enough unique features of the event to accurately classify it. However, as described in the survey comparison section, this phenomenon does not diminish the extremely strong correlation between detected proximity and self-report interactions. Lastly, as discussed in the relationship inference section, while frequency of proximity within the workplace can be useful, the most salient data comes from detecting a proximity event outside MIT, where temporarily forgetting the phone is less likely to repeatedly occur.

Missing Data. Because we know when each subject began the study, as well as the dates that have been logged, we can know exactly when we are missing data. This missing data is due to two main errors discussed above: data corruption and powered-off devices. On average we have logs accounting for approximately 85.3% of the time since the phones have been deployed. Less than 5% of this is due to data corruption, while the majority of the missing 14.7% is due to almost one fifth of the subjects turning off their phones at night.

Surveys & Diaries vs. Phone Data. In return for the use of the Nokia 6600 phones, students have been asked to fill out web-based surveys regarding their social activities and the people they interact with throughout the day. Comparison of the logs with survey data has given us insight into our dataset's ability to accurately map social network dynamics. Through surveys of approximately forty senior students, we have validated that the reported frequency of (self-report) interaction is strongly correlated with the number of logged BTIDs ($R=.78$, $p=.003$), and that the dyadic self-report data has a similar correlation with the dyadic proximity data ($R=.74$, $p<.0001$).[3] Additionally, a subset of subjects kept detailed activity diaries over several months. Comparisons revealed no systematic errors with respect to proximity and location, except for omissions due to the phone being turned off.

[3] Interestingly, the surveys were not significantly correlated with the proximity logs of the incoming students. This phenomena will be addressed in a later paper (Eagle, Lazer, and Pentland, 2005) discussing the fallibility of self-report data in particular situations.

4 Community Structure: Complex Social Systems

In the previous section we showed that Bluetooth-enabled mobile phones may be used to discover a great deal about the user's patterns of activity. In this section we will extend this base of user modeling to explore modeling complex social systems.

By continually logging and time-stamping information about a user's activity, location, *and proximity to other users*, the large-scale dynamics of collective human behavior can be analyzed. If deployed within a group of people working closely together, correlations between the phone log and proximity log could also be used to provide insight behind the factors driving mobile phone use. Furthermore, a dataset providing the proximity patterns and relationships within large groups of people has implications within the computational epidemiology communities, and may help build more accurate models of airborne pathogen dissemination, as well as other more innocuous contagions, such as the flow of information.

Fig 9. Movement and communication visualization of subjects around cell towers.

4.1 Human Landmarks

As shown in Figure 4 and 11, there are people who users only see in a specific context (in this instance, at work). If we know the user is at work, information about the time of day, and optionally the location within the building (using static Bluetooth devices) can be used to calculate the probability of that user seeing a specific individual, by the straightforward application of Bayes' rule.

In contrast to previous work that requires access to calendar applications for automatic scheduling [16], we can generate inferences about whether a person will be seen within the hour, given the user's current context, with accuracies of up to 90% for 'low entropy' subjects. These predictions can inform the user of the most likely time and place to find specific colleagues or friends. We believe that the ability to reliably instigate casual meetings would be of significant value in the workplace. We must also remember, however, that the ability to predict people's movements can be put to less savory uses. Careful consideration must be given to these possibilities before providing free access to such data.

4.2 Relationship Inference

In section 3 we discussed how information about location and proximity can be used to infer a user's context. In much the same way, knowledge of the shared context of two users can provide insight into the nature of their association. For example, being near someone at 3pm by the coffee machines confers different meaning than being near them at 11pm at a local bar. However, even simply proximity patterns provide an indication of the structure of the underlying friendship network as shown in Figure 10. The clique on the top left of each network are the Sloan business students while the Media Lab senior students are at the center of the clique on the bottom right. The first year Media Lab students can be found on the periphery of both graphs.

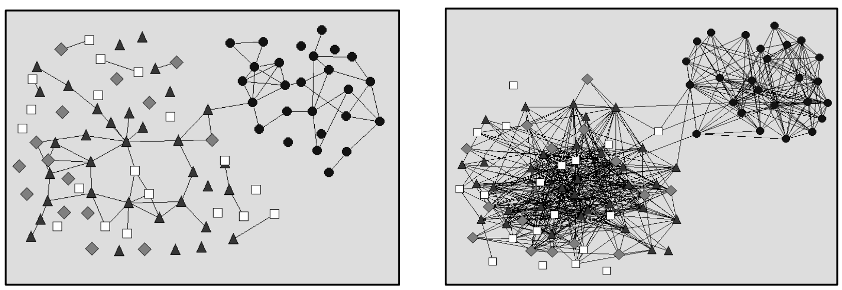

Fig 10. Friendship (left) and daily proximity (right) networks share similar structure. Circles represent incoming Sloan business school students. Triangles, diamonds and squares represent senior students, incoming students, and faculty/staff/freshman at the Media Lab.

We have trained a Gaussian mixture model [5] to detect patterns in proximity between users and correlate them with the type of relationship. The labels for this model came from a survey taken by all of the experimental subjects at the end of two months of data collection (some users came late to the study, but were included anyway). The survey asked who they spent time with, both in the workplace and out of the workplace, and who they would consider to be in their circle of friends. We compared these labels with estimated location (using cell tower distribution and static Bluetooth device distribution), proximity (measured from Bluetooth logs), and time of day.

Workplace colleagues, outside friends, and people within a user's circle of friends were identified with over 90% accuracy, calculated over the 2000 potential dyads. Initial examination of the errors indicates that the inclusion of communication logs combined with a more powerful modeling technique, such as Support Vector Machine, will have considerably greater accuracy.

Some of the information that permits inference of friendship is illustrated in Figure 11. This figure shows that our sensing technique is picking up the common-sense phenomenon that office acquaintances are frequently seen in the workplace, but rarely outside the workplace. Conversely, friends are often seen outside of the workplace, even if they are co-workers. Determining membership in the 'circle of friends' requires cross-referencing between friends: is this person a member of a cluster in the out-of-office proximity data?

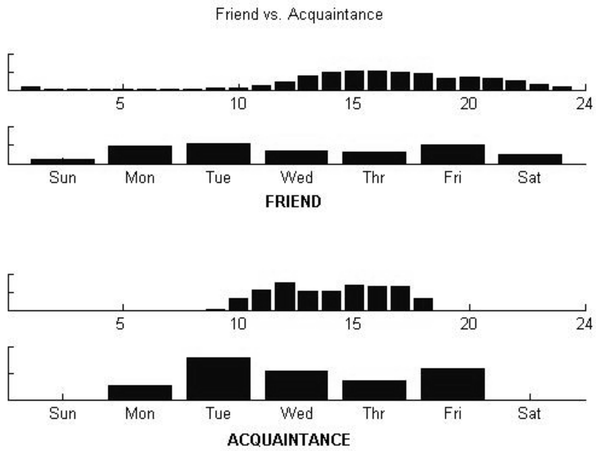

Fig 11. Plotted is proximity frequency data for a friend and a workplace acquaintance of one subject.

	Friends		Not Friends	
	avg	std	avg	std
Total Proximity (minutes / day)	72	150	9.5	36
Saturday Night Proximity (minutes / week)	7.3	18	.20	1.7
Proximity with no Signal (minutes / day)	12	20	2.9	20
Total Number of Towers Together	20	36	3.5	4.4
Proximity at Home (minutes / day)	3.7	8.4	.32	2.2
Phone Calls / day	.11	.27	.001	.017

Table 1. Statistics correlated ($.25<R<.8$, $p<.001$) with friendship generated from sixty subjects (comprising 75 friendships) who work together at the Media Lab

4.3 Proximity Networks of Work Groups

By continuously logging the people proximate to an individual, we are able to quantify a variety of properties about the individual's work group. Although most work in networks assumes a static topology, proximity network data is extremely dynamic and sparse. We are currently building generative models to attempt to paramerterize the underlying dynamics of these networks to gain insight into the functionality of the group itself. Additionally, we hope that quantifying these proximity networks and contrasting the dynamics of the different groups at the Media Lab, we will gain some insight into the underlying characteristics of the research groups.

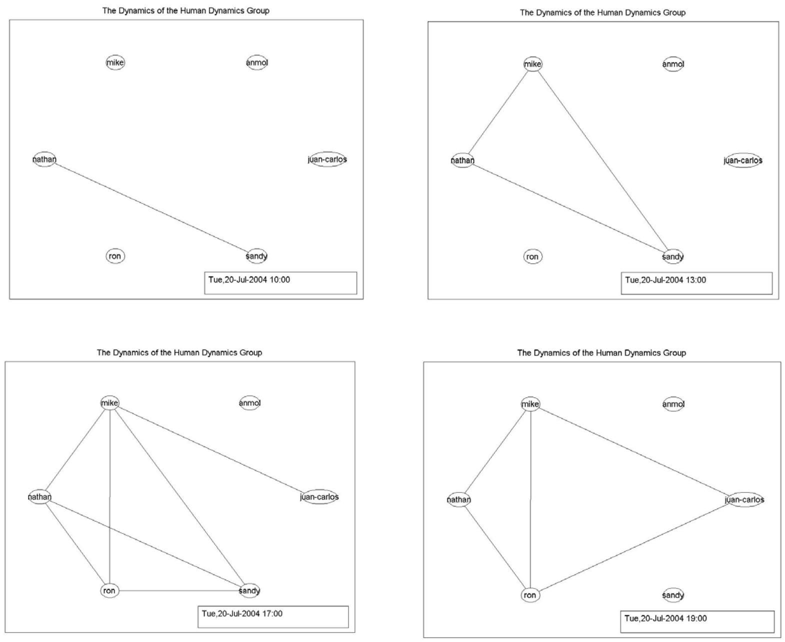

Fig 12. Proximity Networks for a team over one day

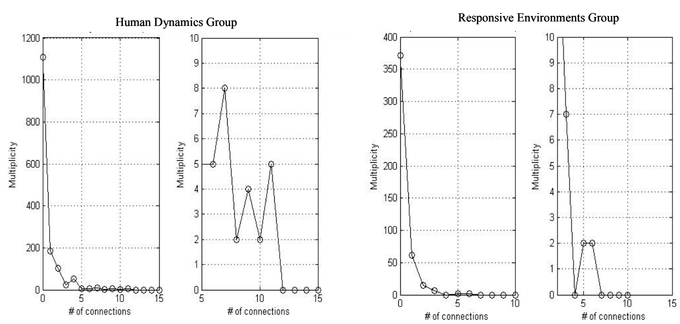

Fig 13. Frequency of Intra-Group Connections. The distribution for both groups has decay factor of approximately -1.5.

4.4 Organizational Rhythms and Network Dynamics

Organizations have been considered microcosms of society, each with their own cultures and values [21]. Similar to society, organizational behavior often shows recurrent patterns despite being the sum of the idiosyncratic behavior of individuals [4]. We are beginning to explore the dynamics of behavior in organizations in response to both external (stock market performance, a Red Sox World Series victory) and internal (deadlines, reorganization) stimuli.

During October, the seventy-five Media Lab subjects had been working towards the annual visit of the laboratory's sponsors. Preparation for the upcoming events typically consumes most people's free time and schedules shift dramatically to meet deadlines and project goals. It has been observed that a significant fraction of the community tends to spend much of the night in the lab finishing up last minute details just before the event. We are beginning to uncover and model how the aggregate work cycles expand in reaction to these types of global deadlines. Figure 14 is a time series of the maximum number of links in the Media Lab proximity network during every one hour window. It can be seen that the number of links in the Media Lab proximity network remained significantly greater than zero during the third week of October and in early December, representing preparation for a large Media Lab sponsor event and MIT's finals week. A Fourier transform (Figure 14, bottom) of this times series uncovers two fundamental frequencies, the strongest being at 24 hours (1 day), and the second being at 168 hours (7 days).

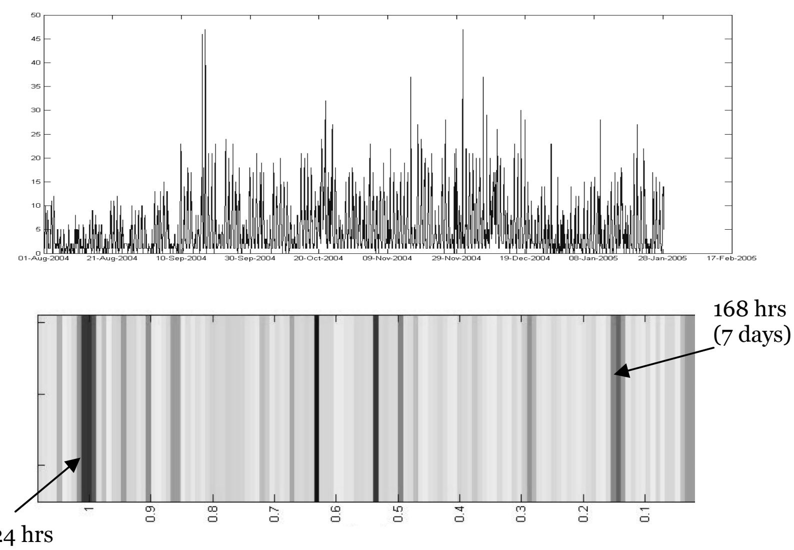

Fig 14. The total number of edges each hour in the Media Lab proximity network from August 2004 to January 2005. Below is its corresponding Fourier transform confirming the two most fundamental frequencies of the dynamic network to be (not surprisingly) 1 day and 7 days.

5 Conclusions

It is inevitable that mobile devices of tomorrow will become both more powerful and more curious about their user and his or her context. We have distributed a fleet of one hundred curious mobile phones throughout a laboratory and a business school at MIT. The data these devices have returned to us is unprecedented in both magnitude and depth. The applications we have presented include ethnographic studies of devices usage, relationship inference, individual behavior modeling and group behavior analysis. However, there is much more to be done, and it is our hope that this new type of data will inspire research in a variety of fields ranging from qualitative social science to theoretical artificial intelligence.

Acknowledgements

The authors would like to express their gratitude to Wen Dong, Stephen Guerin, Tony Pryor and Aaron Clauset. We would also like to thank Hari Pennanen and Nokia for their supporting. Finally, Mika Raento deserves particular recognition as the architect of Context and whose efforts were instrumental to this research.

References

1. Ashbrook D, Starner T, "Using GPS to learn significant locations and predict movement across multiple users", Personal & Ubiquitous Computing (2003) 7: 275-286.

2. Bar-Noy A, Kessler I, "Tracking mobile users in wireless communication networks," IEEE Transactions on Information Theory, 39(6): 1877-1886, November 1993.

3. Bhattacharya A, Das SK, "LeZi-update: an information-theoretic approach to track mobile users in PCS networks. In: Proceedings of the International Conference on Mobile Computing and Networking, Seattle, WA, August 1999.

4. Begole, J., Tang, J., Hill, R., "Rhythm modeling, visualizations and applications", Proceedings of the UIST 2003. p 11-20.

5. Duda Ro, Hart PE and Stork DG, "Pattern classification" Wiley, New York 2001.

6. Eagle, N., and Pentland, A., "Mobile Matchmaking: Proximity Sensing and Cuing"', IEEE Pervasive Computing. To appear: April 2005.

7. Himberg J, Korpiaho K, Mannila H, Tikanmäki J, and Toivonen H.T.T. "Time series segmentation for context recognition in mobile devices". In: Proceedings of the IEEE International Conference on Data Mining (ICDM 2001), pp. 203-210, San José, California, USA, 2001.

8. Kim SJ, and Lee CY, "Modeling and analysis of the dynamic location registration and paging in microcellular systems," IEEE Transactions on Vehicular Technology, 45(1):82-90, February 1996.

9. Laasonen K, Raento M, Toivonen H, "Adaptive On-Device Location Recognition", In: Proceedings for Pervasive, pp 287-304, 2004.

10. Laibowitz, M., "Parasitic Mobility for Sensate Media", MS Thesis in Media Arts and Sciences. Cambridge, MIT (2004).

11. LaMarca, A. et al., "Place Lab: Device Positioning Using Radio Beacons in the Wild". To appear, Pervasive 2005, Munich Germany.

12. Liao L, Fox D, Kautz H, "Learning and Inferring Transportation Routines" In: Proceedings for the National Conference on Artificial Intelligence (AAAI-04), San Jose, CA, July 2004.

13. Mäntyjärvi J, Himberg J, Kangas P, Tuomela U, Huuskonen P, "Sensor signal data set for exploring context recognition of mobile devices" In: Workshop "Benchmarks and a database for context recognition" in conjunction with the 2nd Int. Conf. on Pervasive Computing (PERVASIVE 2004), April 18-23, Linz/Vienna, Austria, 2004.

14. Paulos E, Goodman E, "The familiar stranger: anxiety, comfort, and play in public places". CHI 2004 223-230.

15. Raento, M., Oulasvirta, A., Petit, R., Toivonen, H., "ContextPhone – A prototyping platform for context-aware mobile applications". IEEE Pervasive Computer. To appear: April 2005.

16. Roth J, Unger C, "Using handheld devices in synchronous collaborative scenarios" In: Proceedings of the Second International Symposium on Handheld and Ubiquitous Computing (HUC), Bristol, UK, 25-27 September 2000.

17. Singh, P. and Williams, W., "LifeNet: a propositional model of ordinary human activity". Proceedings of the Workshop on Distributed and Collaborative Knowledge Capture (DC-KCAP) at K-CAP 2003.

18. Want, R., Hopper, A., Falcao, V., and Gibbons, J., "The active badge location system," ACM Transactions on Information Systems, vol. 10, pp. 91--102, Jan. 1992.

19. Wolf J, Guensler R., and Bachman W., "Elimination of the travel diary: an experiment to derive trip purpose from GPS travel data". In: Proceedings from the Transportation Research Board 80th annual meeting, Washington, DC, 7-11 January 2001

20. Weld D, Anderson C, Domingos P, Etzioni O, Gajos K, Lau T, Wolfman S, "Automatically Personalizing User Interfaces", In: Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI03), Acapulco, Mexcio, 2003.

21. Wertheim E, "Historical Background of Organizational Behavior" http://www.business.com/directory/management/management_theory/organization_behavior_and_culture/

22. http://news.bbc.co.uk/1/hi/business/4257739.stm

23. http://www.bluetoothsig.com/emailfiles/Zelos.pdf