Nodobo: Mobile Phone as a Software Sensor for Social Network Research

Stephen Bell, Alisdair McDiarmid, and James Irvine
Centre for Intelligent and Dynamic Communications
Dept. of Electronic and Electrical Engineering
University of Strathclyde
Glasgow, UK
G1 1XW
Email: {stephen.bell,a.mcdiarmid,j.m.irvine}@strath.ac.uk

***Abstract*—Modern smart phones are now capable of gathering information about a user's social interactions. The authors have developed and deployed Nodobo, a suite of social sensor software for Android. Our first study group is a class of senior high school students, each using a Google Nexus One mobile phone running Nodobo, which we use to capture their device usage patterns and social interactions. We provide an overview of the system architecture, describe the trial, and share some initial results.**

I. Introduction

Mobile phones are now ubiquitous across all sectors of society. The permanent presence of the phone in your pocket allows it to act as a social sensor, detecting interactions with the people around you every day. A recently-started project at the University of Strathclyde is using this technology to automatically detect the social networks within a group of graduating high school students. This data will then be analysed to determine the impact of peer support on effective learning at this crucial stage of the education process.

II. Rationale

The continuous growth of the mobile handset sector over the past decade has led to an enormous increase in the level of technology available at the top end of the handset market. Modern smart phones are better described as powerful handheld computers with permanent Internet connections. This power is largely untapped, with the phone idle for the majority of the day. Our project uses these resources to record social interactions within a group, generating a dataset with which we can infer social links and describe an overall social network[1].

Reality Mining[2] is the first detailed study of communicative ties between mobile phone users with publicly-available results. The experiment involved approximately 100 mobile phone users in a university setting, monitoring phone calls, text messages, and Bluetooth proximity. The resulting data set is generously made available upon request from the authors, but is small, incomplete, and we have found some concerning internal inconsistencies[3].

One aim of this study is to capture a broader set of data, with greater robustness and reliability, from participants who use their devices more often. Once completed, we hope to anonymise and release the dataset to the research community.

Another goal is to examine how interactions among peer groups affect the learning outcomes of high school students. The study group is regularly interviewed to verify the usage of the phones is accurate. At the end of the study we will be able to look for correlation between academic performance and social interactions.

III. Architecture

Our researchers developed "Nodobo", a set of software extensions to the Google Android operating system, for enabling the capture of social context. Since Android is almost completely open source, source code modifications facilitate the capture of the required metadata. The modifications extend existing communications applications included with Android: no custom applications were required. The system architecture is outlined in Figure 1.

The software captures a variety of social context data, including logs of phone calls, text messages, Bluetooth proximity detection, WiFi access point, and cell tower ID. The directionality of calls and text messages are recorded, along with the associated phone number, and the duration of the call or length of the message. Bluetooth proximity is detected every minute, and includes all devices in the study as well as any other clients which respond to service discovery. Basic positioning is achieved through WiFi hotspot and cell tower ID records.

While this project is examining the private communications data of its users, we are taking steps to limit the level of privacy invasion as much as possible while achieving our research goals. The audio of phone calls is not inspected or recorded, and the content of SMS is only used to determine the length of the messages. Data are stored securely on a password protected server. Only the senior research co-ordinator has the record which links the phone IMEI to the individual using the device.

Social interactions are captured to a flat database on the device's SD card. A flat database is used to make it easier to add extra context sensors as required: using a normalised database would require a new schema to be deployed with each new sensor. The system uses a sensor/generator approach commonly found in context-gathering systems[4], [5].

Past calls and SMS messages are stored in their own respective databases called 'contacts.db' and 'mmssms.db'.

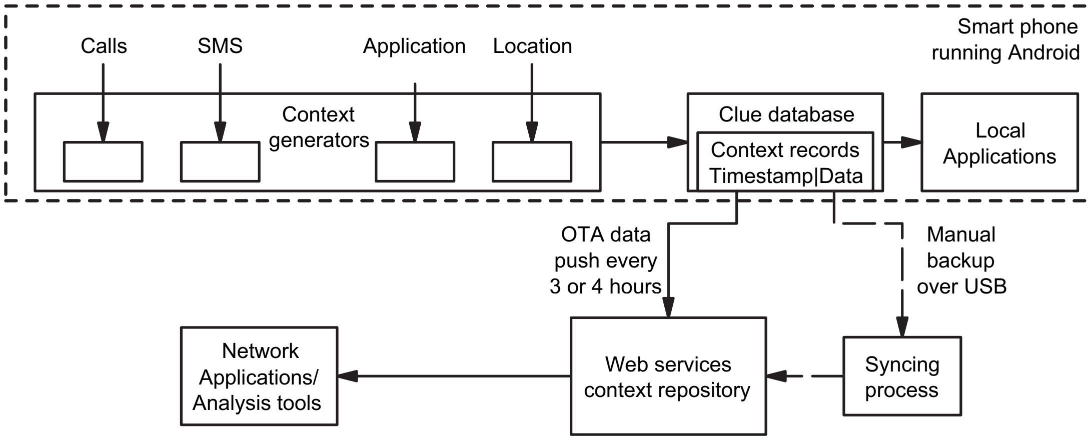


Fig. 1: Nodobo architecture diagram

Android provides a mechanism for obtaining information from these databases through objects called 'content providers' and 'content observers'; specifically, an application can register to receive notifications when the content of a database changes. By registering a content observer for each of the databases, it is possible to record when calls and SMS messages are sent and received.

Multiple notifications are sent each time a call or SMS event occurs, so a system which takes the latest entry in the database would lead to the same call or message being stored multiple times. To get around this, the notification handler first fetches the latest original ID from the social store, then selects all interactions greater than this value from the calls database. These interactions are then added to the social store. Therefore, if multiple notifications are called for the same interaction event, only the first will trigger the event to be added to the social store.

For recording proximity interactions, the Bluetooth radio is used to detect nearby devices. The software enters discovery mode every minute, and attempts to discover devices for 12 seconds. As the Bluetooth modem is unavailable when the device is asleep, the application must wake the device in order to discover devices. An alarm is registered, which wakes the device every polling period and registers a wake lock so the processor will not go back to sleep once the discovery process is started. When a Bluetooth device is detected, a callback is used to store the proximity interaction in the database. Upon completion of discovery, the software releases the wake lock, allowing the processor to go back to sleep, saving battery life.

IV. Redundancy

The phone database is synchronised periodically over-the-air with a web services data store. With this approach, we continuously receive data from the devices throughout the study, can detect broken or out of contact devices quickly, and still have the backup solution of retrieving the SD card at the end of the project.

Three backup strategies were considered: using the mobile networks for backups, installing a WiFi network at the location of the deployment, and using existing WiFi networks that the devices may connect to. The strategies vary in terms of financial cost to the organisers and participants, effort, and reliability (see Table I).

From a financial perspective, having the devices backup their data using mobile networks or the user's WiFi connection doesn't incur any extra cost to the organisers. Installing a standalone WiFi network at the deployment location would require some expense. A backup server could be deployed on-location, and connected to the same WiFi network as the devices. If the backup server cannot be connected to the internet through networks available at the deployment, a 3G modem could be used. The backup server could then communicate with an off-site web server for redundancy.

If backups were to be made over the mobile networks, the user would require a data plan, which may incur a fixed cost per month. There would be no cost in using a wifi network deployed by the study organisers, and wifi is becoming very popular in the home, so this probably wouldn't incur any cost to the user either.

In terms of effort, using the mobile network or user's wifi connections doesn't require any additional effort on the part of the organisers. However, deploying a wireless network specifically for study use requires the organisers to travel to the deployment and install and configure a wireless network. From a user's effort perspective, if they don't have a data plan, then they need to get one added to their contract or PAYG plan. They would also have to either manually connect to any WiFi networks, which they can't be relied upon to do. Most devices connect to remembered wireless networks automatically, if the WiFi is turned on, however, this has a detrimental effect on battery life.

Deploying a wireless network will only be successful if the study was cohesive, and all participants were expected to pass through the same place at one point or another (e.g., a workplace, school, or university). If the study is distributed over a number of different locations, then to facilitate a wireless network deployment, an access point would need to be provided for each location, increasing cost.

	Mobile Network	Our Wifi	Their Wifi
Deployer Cost	No cost	Cost of hardware and installation	No cost
User Cost	Cost of data plan	No cost	Marginal cost
Deployer Effort	No effort	Installation of Wifi	No effort
Coverage	Nationwide	Single area	Single area
Connectivity	Automatic	User-driven	User-driven

TABLE I: Redundancy summary

V. METHODOLOGY

A group of 27 promising high school students in a Scottish state high school were selected for this study. The participants are in their 5th year of high school, undertaking Higher qualifications with an aim to achieving a university place in one or two years' time. All students previously had a mobile phone, with approximately 1/3 of these falling in the category of smartphone (iPhone, Blackberry, or similarly powerful handset).

The close proximity of the deployment to the university enables the study organisers to schedule regular visits to diagnose issues, as well as facilitating regular backups phone backups to be made. To maintain as up-to-date a dataset as possible, and to limit the number of visits required, the devices also synchronise with a web server over the mobile network or WiFi.

VI. APPLICATION

Social networking tools for mobile devices enable peer support of learning in new ways. Less conspicuous than more traditional peer support mechanisms such as study groups, there are interesting social consequences in allowing interaction where the dominant peer pressure may be not to study. In addition, new services like activity indicators (showing applications used by individuals, or where they are located) have the potential for changing the peer support dynamic in ways which have not yet been studied.

With our detailed data, gathered reliably and securely, many opportunities for later analysis arise. The immediate aim of the project will be to analyse the gathered results to detect social ties and relationships based on interactions between users. This information will be compared with survey results and cross-referenced with other data to determine the relationship between social interactions and educational performance.

Our long-term aim is to examine the use of social-support mobile phone applications which can foster valuable ties between people. This will work towards reducing isolation, increasing peer support, and improving academic performance.

VII. RESULTS

Four months into the study, around 4.5 million interactions have been generated, including 11733 phone calls, 107454 SMS messages, 4289127 proximity interactions, 113447 cell tower records, and 24476 WiFi access point detections. Despite using around a quarter of the number of participants, this is out-pacing the Reality Mining study data significantly: Reality Mining included 18760 calls, 3590 SMS messages, and 285512 proximity interactions over the full year-long study.

Figure 2 shows an examination of calls, SMS messages, and proximity interactions broken down by the hour of day each occurred. As would be expected, the vast majority of proximity interactions occur during the school day: from 8am to 3pm. Despite the school policy banning the use of phones during lesson time, we see a significant number of calls and SMS also occurring during this period, but both of these increase significantly after school and throughout the evening. Perhaps also of note is that phone calls rapidly drop off around 10pm, but SMS messages continue to be sent until after 2am.

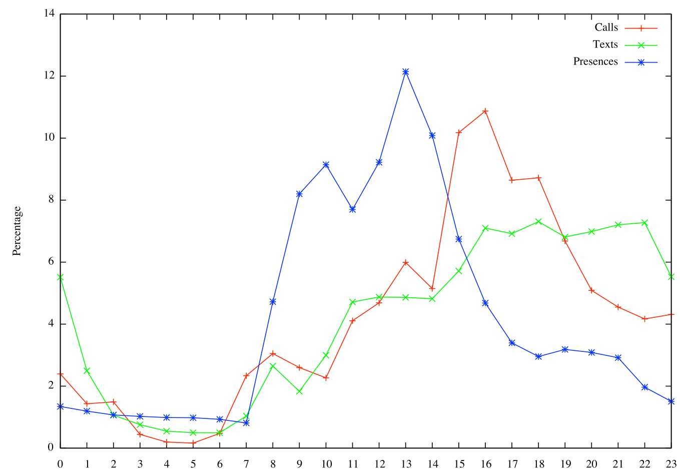


Fig. 2: Calls, SMS, and proximity detections, grouped by hour.

One of the main purposes of gathering communications metadata is to examine the concept of computing social graphs as exhibited by the use of mobile phones[6]. We examined communications among the study group, and plotted social graphs based upon the existence of these links. Figure 3a and Figure 3b show the social graphs produced by examining call and message interactions. Each node represents a numbered study participant, with the background colour indicating betweenness centrality, ranging from blue (most central) to red (least central). Betweenness centrality[7] is a measure of how many shortest paths on a graph flow through a given node; in a social network, it can measure how important a person is to communications across the group.

An edge exists between two nodes if they have dichotomous communications, i.e. if both have initiated the call or message. For this group of participants, calls are very nearly a subset of SMS, with only the edge 4–13 being calls-only. For this young study group, SMS traffic is more than an order of magnitude greater than call traffic, so this result is not unexpected.

Deriving meaning from the presence data is challenging, due to its large quantity and high degree of variation among users. Presence is by definition reciprocal, and a graph of all co-presence ties is almost fully connected, as almost all of the users have been in physical proximity at least once. Our initial attempt at solving this problem is to identify links representing "regular presence", which is arbitrarily defined as at least 30 minutes spent together on 75% of days in the study. We only consider days where both users have detected some presence, to ignore periods where one or both phones are off. The resulting social graph is shown in Figure 3c.

Figure 4 shows the combination of the networks in Figure 3. We can see that the larger group is linked to the smaller one by presence, a form of non-mediated communication. This is notable as it suggests that capturing mediated communications (that is, calls and texts) is insufficient to build an accurate social graph. Presence forms important bridges between the two main groups, with the relationships between 1–16–23, with user 1 acting as a vital router or broker[8].

The combination of calls, texts, and proximity interactions still leaves isolated nodes. However, it is important to remember that the study group is a subset of a larger social network in the school. These nodes may be isolated within the scope of the study, but they do have many communications with other people outside the study.

It is notable that regular presence estimated by this function establishes multiple spanning links. For example, 1–(19, 21), and 16–(1, 23): presence each time forming two links to link nodes together, rather than just a single link between people, as with calls and messages. Initial examinations of the raw data suggest that this is due to socialising in small groups, in which proximity indicates a different kind of relationship status than direct communications.

VIII. Summary and Future Work

By deploying our Nodobo social sensor software with the Google Nexus One mobile phone, we are studying the social interactions of a group of senior high school students. Data is collected from the phone logs, synchronised regularly with a web services data store, and can be examined and studied to determine how the study participants use the devices to foster and maintain social links among their peers. At the current rate of collection, our data set is expected to be significantly larger and more detailed than the current best public data set, Reality Mining.

Examination of the gathered data to produce results similar to those shown in Figure 2 is just one possible direction for exploration. As the study progresses, we will begin more detailed analyses of the social information which can be derived from these data. We will develop and test techniques for estimating and valuing social links between the study participants, and validate our results through qualitative and quantitative methods. Our goal is to extend this work towards developing applications for smart phones which facilitate and support peer support to improve the educational outcomes for future learners.

Most importantly, this study shows that the openness and available processing power of modern smart phones makes it feasible to deploy a system to capture social context. The pervasiveness of the phone ensures that a rich dataset can be generated quickly, using tools which create minimal intrusion to user experience. Our software captures vast quantities of data associated with a user's social context, which can then be used immediately for research, or in the future to support advanced ubiquitous mobile applications.

References

[1] Stephen Bell, Alisdair McDiarmid, Jamie Banford, and James Irvine. Nodobo social: Using smartphones for social science. Wireless World Research Forum Meeting 25, November 2010.

[2] Nathan Eagle and Alex (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10:255–268, March 2006.

[3] Jamie Banford, Alisdair McDiarmid, and James Irvine. Estimating the Strength of Ties in Communication Networks with a Small Number of Users. In *Proceedings of The Sixth International Conference on Wireless and Mobile Communications*, September 2010.

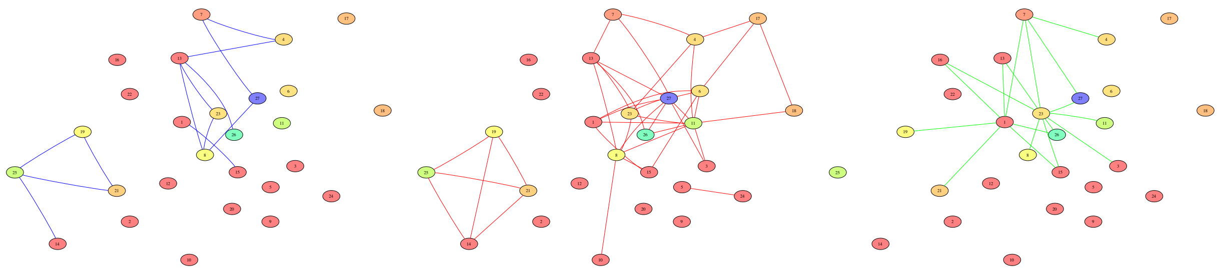
[4] Patrick Fahy and Siobhan Clarke. Cass – a middleware for mobile context-aware applications. In *Workshop on Context Awareness, MobiSys*, 2004.

[5] Daniel Salber, Anind K. Dey, and Gregory D. Abowd. The context toolkit: aiding the development of context-enabled applications. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, CHI '99, pages 434–441, New York, NY, USA, 1999. ACM.

[6] Jamie Banford, Alisdair McDiarmid, and James Irvine. Multigraph Representation of Relationships for Enterprise Knowledge Networks. Wireless World Research Forum Meeting 22, May 2009.

[7] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[8] Valdis Krebs. The social life of routers: Applying knowledge of human networks to the design of computer networks, 2000.



(a) Calls (b) Texts (c) Proximity

Fig. 3: Dichotomous interactions by type

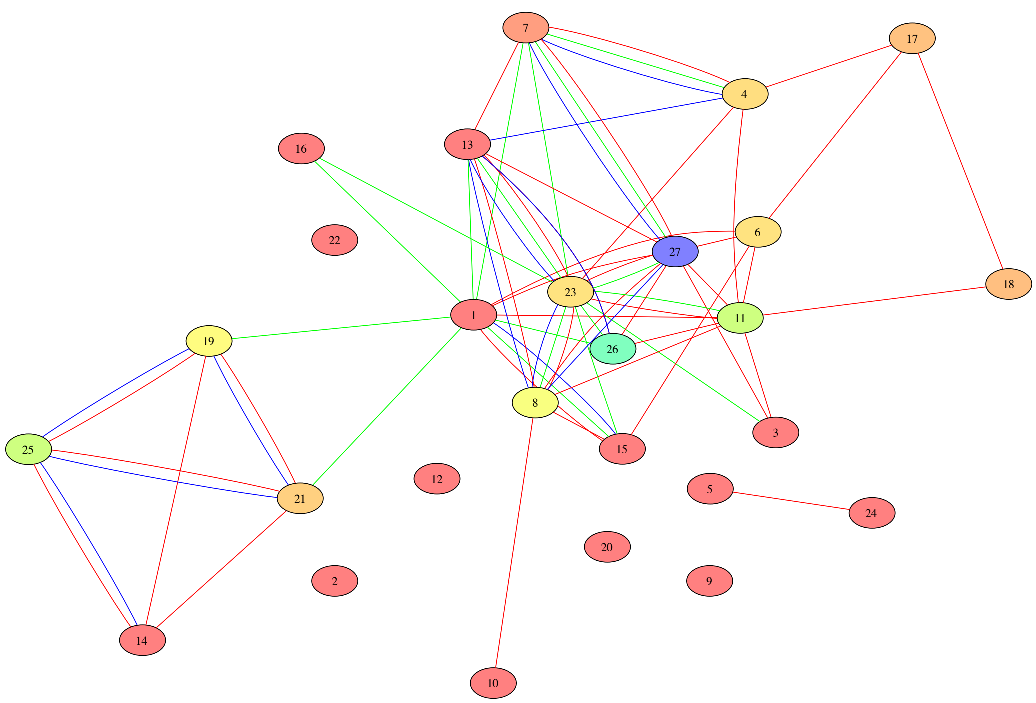


Fig. 4: Cumulative tie data