Nodobo: Detailed Mobile Phone Usage Dataset

Alisdair McDiarmid, Stephen Bell, James Irvine, Jamie Banford
Department of Electronic & Electrical Engineering,
University of Strathclyde,
Glasgow, United Kingdom
Email: a.mcdiarmid@strath.ac.uk

Abstract

We announce a mobile phone usage dataset, gathered through monitoring devices of 27 users over a 5-month study. The data includes 13035 call records, 83542 SMS records, and 5.2 million proximity records. Each datum has a high level of detail, and the combination of multiple kinds of data makes this a valuable resource for a variety of research topics. We include some study results, and describe how to access and use the dataset.

1 Introduction

The ubiquitous mobile phone creates the possibility of automatically detecting tie signs—the subtle ways in which people reveal the nature of their relationships[1]. As part of a study of the social network established in a group of 27 high school students, we have captured a set of data which can be mined for these tie signs. We describe the data, and some initial results. The dataset itself has been published for use by the research community.

1.1 Related Work

There have been several investigations of mobile phone usage as it relates to social ties. Onnela et al. showed that call duration can be used as a proxy for tie strength [2]. Lambiotte examined the relationship between geographic proximity and social network distance[3]. These studies use

private datasets from mobile operators; researchers without access to similar data cannot reproduce their results or test their own theories.

Reality Mining[4] is the first detailed study of communications and proximity between mobile phone users with a published dataset. The experiment involved approximately 100 mobile phone users in a university setting, monitored over an academic year. The resulting data set is made available upon request from the authors, but it represents usage patterns which are more than five years old, and the data suffers from sparsity of mediated communications[5]. This is the motivation for the release of our new *Nodobo* dataset, which directly addresses these issues.

2 Data Collection

Each of the study participants was given a Google Nexus One smartphone, prepared with a modified Android operating system. The openness of the Android platform allowed our researchers to transparently capture information about phone usage, without any effect on the user experience. Data is stored in a simple database on the device SD card, which is then synchronised over the air to a central server[6].

The software on the phone captures data using a variety of software sensors, logging phone calls and text messages, Bluetooth device discovery, WiFi access point, and cell tower ID. The direction of calls and text messages is recorded, along with the associated phone number, and the duration of the call or length of the message. Bluetooth proximity is recorded every minute, and includes all phones in the study as well as any other devices which respond to service discovery. Basic positioning is achieved through WiFi hotspot and cell tower ID records.

3 Results

From September 2010 to the end of January 2011, our study recorded 13035 call records, 83542 SMS records, and 5.2 million proximity records. While the main contribution of this paper is publishing the data itself, we also present three short studies with results.

Figure 1 shows an examination of calls, SMS messages, and proximity interactions grouped by hour of day. As expected, the majority of proximity interactions occur during the school day, from

8am to 3pm. A significant number of calls and SMS also occur during this period, but this activity increases significantly after school and throughout the evening. Note that phone calls rapidly drop off around 10pm, but SMS messages continue to be sent until after 2am.

In one previous study, most SMS conversations were found to be only two messages long, with the number of conversations decreasing rapidly as the number of messages increases[7]. Figure 2 shows a comparison between the results published in [7] and those derived from our dataset. The original study examined semantic meaning to form conversations, but we do not have access to message contents. Our results group messages based upon communicants, direction, and a windowed reply time between messages. Results are given for window sizes $w = 120$ and $w = 180$ seconds, and show strong similarity to the previous results.

Monitoring both communications and proximity allows us to estimate the study group's social network by combining these two different types of data[8]. Figure 3 shows a multigraph, where each node represents a person, and each edge represents significant communications or proximity. A solid blue edge exists between two people if they have exchanged calls or text messages in both directions. A dashed green edge exists when the two people have spent at least 30 minutes per day in proximity to each other, for at least 4 days per week. These dichotomous measures are intended to identify significant ties between the participants. The resulting data-derived graph matches closely to the actual social network, as captured using qualitative methods.

There are clearly two major social groups, with some participants being more centrally-connected than others. We can see that the larger group is linked to the smaller one by proximity alone, and that there is significant proximity interactions only within the larger group. This indicates that considering direct communication (calls, SMS, email) is insufficient to build a fully rich social graph.

3.1 Data Availability

This dataset has been anonymised and published at our project website, Nodobo.com. The free release includes basic modelling tools to enable rapid investigations of the data, as well as the source code required to reproduce the results given here. The Nodobo trial continues throughout 2011, and we intend to update the dataset as we gather more results.

4 Summary

We present the Nodobo dataset, a public release of mobile phone usage data. This data includes call, SMS, and proximity records of 27 participants over a 5-month period. Several examinations of the data have reproduced results from other studies, and given insight into how the participants use their mobile phones. We have released the data and supporting tools in the hope that it is useful to other researchers in this area.

More broadly, this study shows that the openness of modern smart phone platforms makes it feasible to deploy a system to capture social context. The pervasiveness of the mobile 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] M. A. Smith, *The Hyperlinked Society: Questioning Connections in the Digital Age*, ch. From Hyperlinks to Hyperties, pp. 165–176. digitalculturebooks, 2009.

[2] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and Tie Strengths in Mobile Communication Networks," *Proceedings of the National Academy of Sciences*, vol. 104, pp. 7332–7336, May 2007.

[3] R. Lambiotte, V. D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. V. Dooren, "Geographical dispersal of mobile communication networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 21, pp. 5317–5325, 2008.

[4] N. Eagle and A. Pentland, "Reality Mining: Sensing Complex Social Systems," *Personal Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, 2006.

[5] J. Banford, A. McDiarmid, and J. 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.

[6] S. Bell, A. McDiarmid, and J. Irvine, "Nodobo: Mobile phone as a software sensor for social network research," in *Proceedings of Context Awareness for Proactive Systems 2011*, May 2011.

[7] J. Androutsopoulos and G. Schmidt, "Sms-kommunikation: Ethnografische gattungsanalyse am beispiel einer kleingruppe," *Zeitschrift fur Angewandte Linguistik*, vol. 36, pp. 49–80, 2002.

[8] J. Banford, A. McDiarmid, and J. Irvine, "Multigraph Representation of Relationships for Enterprise Knowledge Networks." Wireless World Research Forum Meeting 22, May 2009.

Figure Captions

- Figure 1: Distribution of calls, messages, and presence by hour of day
- Figure 2: SMS conversation length and number of conversations
- Figure 3: Social network map: reciprocal comms and regular presence

Figures

Figure 1

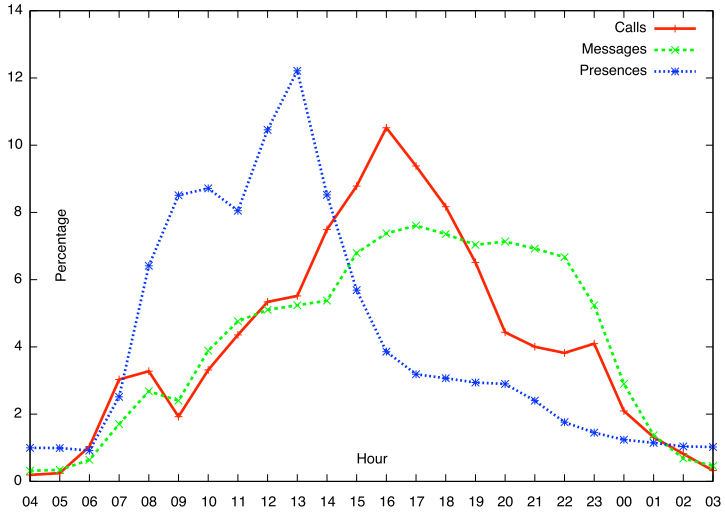


Figure 2

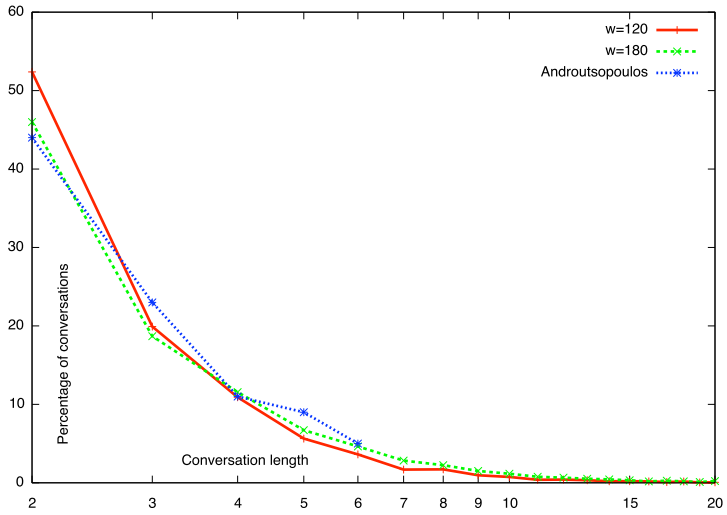
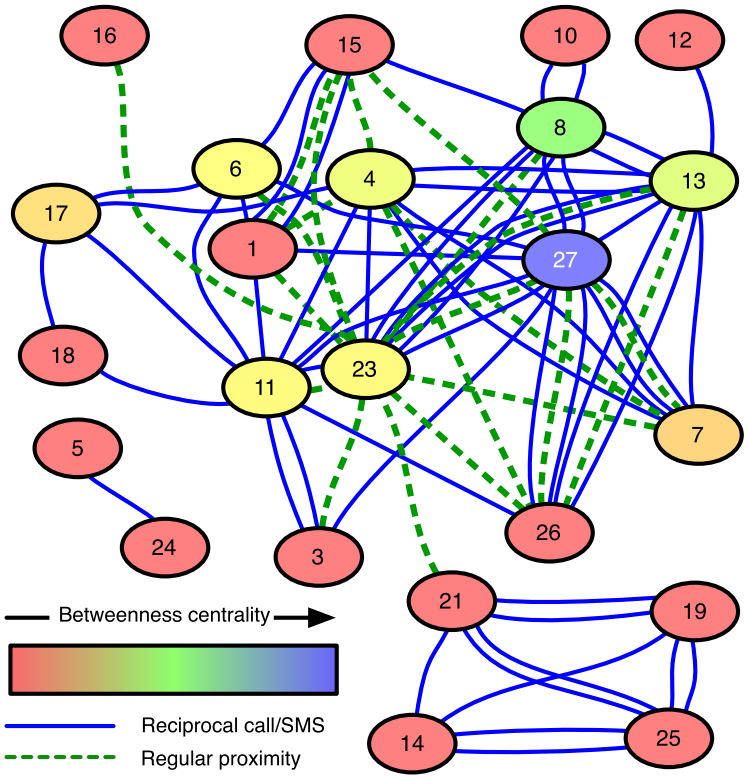

w=120
w=180
Androutsopoulos
Percentage of conversations
Conversation length
60
50
40
30
20
10
0
2
3
4
5
6
7
8
9
10
15
20

Figure 3


16
15
10
12
8
6
4
13
17
1
27
18
23
11
7
5
24
3
26
21
19
14
25
Betweenness centrality
Reciprocal call/SMS
Regular proximity