Tracing Mobile Phone App Installations in the "Friends and Family" Study

Nadav Aharony, Wei Pan, Cory Ip, Alex (Sandy) Pentland

MIT Media Laboratory, Cambridge 02139, MA, USA
{nadav,panwei,coryip,pentland}@media.mit.edu

Abstract

The Friends and Family study in the MIT Media Lab is a long-term mobile phone-based experiment that transforms a graduate family community into a living lab for social science investigation. Data from this study, collected via Android-based phones equipped with our software platform for passive data collection, will be used to look at issues including individual and group identity, real world decision making, social diffusion, social health, and boundaries of privacy. In this paper we give an overview of the study and highlight some of the unique aspects that separate it from previous experiments, and report on some preliminary results from the recently concluded pilot phase. We focus our initial analysis on patterns surrounding mobile applications (apps). We look at participants' app installation patterns and investigate the roles of different networks, inferred from Bluetooth proximity and self-reported surveys, in the spreading of apps. We find that face-to-face interactions have a stronger correlation with the number of shared apps between individuals than self-perceived ties.

Introduction

Today's mobile phones are powerful computing and sensing platforms. We are investigating ways to help users leverage individual as well as aggregated data to improve their lives. Additionally, we are investigating how this data can contribute to the understanding of societal and especially community-related issues.

The Human Dynamics Group in the MIT Media Lab has developed the methodology of Reality Mining, which is defined as the collection and analysis of machine-sensed environmental data pertaining to human social behavior [1] and is a key component in the transformation of traditional social science into the emerging field of computational social science [2]. To gather this data, we use our self-designed sensor platforms as well as smartphones. In recent years, we have performed two large-scale experiments on the MIT campus using close to one hundred phones each. The first study was performed in 2005 with participants from the MIT Media Lab and the Sloan School of Management, who represented a population of Media Lab colleagues and coworkers [1]. The second study was performed during the 2008-2009 academic year at an MIT undergraduate dorm during the 2008-2009 academic year, with a study population comprised of undergraduate students [3].

FunF Study Overview

The Friends and Family study (FunF) is an experiment in the form of a living lab, with participants' everyday behavior patterns sampled via mobile phones and other data collection mechanisms. The pilot phase of the study ran from March to July 2010 with 55 participants, and the expanded second phase of the study will begin in September 2010 with around 200 participants. The data collected pertains to both the physical and digital realms and includes information on face-to-face interactions, mobility, phone communication networks, and online social network activity. The study team also has direct access to the participants in the forms of questionnaires, interviews, and various experimental interventions, giving the FunF study access to a tight-knit physical community at an unprecedented scale and depth. Considering the study will run at least 18 months, the dataset generated from the study will shed light on a wide range of behavioral, social, and health-related topics.

Study Goals

The study touches on many aspects of life, from social dynamics to health to purchasing behavior to community organization. The two high-level topics that unify these varied aspects are: **(a)** how people make decisions, especially the social aspects involved in decision making, and **(b)** how we can empower people to make *better* decisions using personal and social tools.

Study Components

The study is composed four main components:

Android Phone Sensing Platform (FunF System): This is the core of the study's data collection. Android OS-based mobile phones are used as in-situ social sensors to map users' activity features, proximity networks, media consumption, and behavior diffusion patterns. The phones are augmented with our software, which periodically senses and records information such as cell tower IDs; wireless LAN IDs; proximity to nearby phones and other Bluetooth devices; accelerometer and compass data; call and SMS logs; statistics on installed phone applications, running applications, media files, and general phone usage; and other accessible information. The system also supports integration of user-level apps, such as the alarm clock app we developed for additional data collection and potential use in interventions.

Surveys: Each participant has to complete surveys at regular intervals, currently set at weekly and monthly. These include self reports about their perception of their social relationships, groups, and interactions, logging of various types of activities and mood, and standard scales that examine different personality traits and states (e.g. the Big Five Personality Test [4]).

Purchasing Behavior: Information on purchases is collected through receipts and credit card statements submitted at the participants' discretion. This component targets a specific set of categories: child-related, entertainment, and dining expenses.

Facebook Data Collection Platform: Participants can optionally install a Facebook application to log different Facebook activities.

The FunF Community

The study is conducted at a residence hall for married or partnered graduate students at a northeastern US university. The community is composed of over 400 residents, approximately half of which have children of a generally young age, with low- to mid-range household income. The residence has a vibrant community life.

One of the most intriguing of this community is a diverse range of sub-communities and groups. Figure 1 shows an example of a subset of the groups that a single study participant is part of, along with the number of other pilot phase participants in those same groupings.

Study Timeline

During March 2010 a pilot phase of the study was launched with 55 participants. Data was collected for over

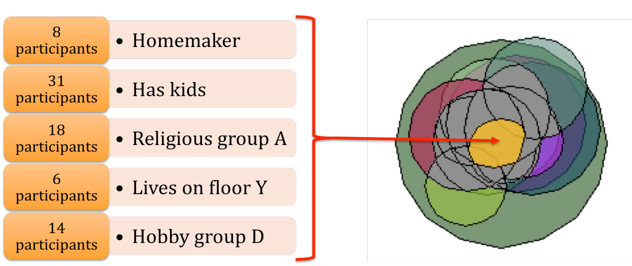

Figure 1: Each circle on the right represents a different social group our participants may belong to. The yellow blob represents one participant, and his affiliated groups along with number of participants in those groups are shown on the left.

three months. The study is aimed to start its second phase in September, expanding to approximately 200 participants. This phase will last for at least a year, and might continue longer dependent on budget constraints and willingness of the community.

Tracing Application Installations

In the remainder of the paper we present for the first time our initial analysis of some of the pilot phase's collected data. In particular, we focus on the topic of mobile apps installed though the Android Market.

Introduction

The mobile device applications industry is booming, and people believe it will grow into a market as big as the Internet [5]. Apps are unique products in many ways: they are easy to build and distribute via the app-store, there are many apps with similar functionality, and the market share change is rapid. Users are overwhelmed by the sheer number of applications available through the market – in July there were over 100,000 apps in the Android Market [6]. Therefore, it remains a challenging and important problem to understand the behavior patterns of individuals purchasing and installing apps and more importantly the spreading mechanism of apps in social groups. App-related research will bring new insights on viral marketing, online marketing, social influence, and information diffusion.

The FunF study provides a unique opportunity to study app spreading mechanisms. The Android platform records every app that a user installs. Overall, in the pilot phase of FunF we collected the app installation activities for all 55 participants from March to early July.

Results

In the pilot study, the 55 participants have installed around 870 unique apps (not counting any apps that come bundled with the phone or the OS version). For this analysis, we only look at app installations and ignore un-installations. We first demonstrate statistics for all of the apps in the study: In Figure 2, we plot the distribution of number of users installing each app. We discover that our data corresponds very well with a power-law distribution with exponential cut. This is normal considering we have a limited number of subjects in this phase. We also plot the distribution of number of apps installed per user in Figure 3, which fits well with an exponential distribution, and suggests that most users will only install a limited number of apps. The implication of this finding is that it is more difficult to promote apps to users if they have already tried many apps previously.

We move on to investigate the network effect of app installations in our study community. To begin with, we look at the proximity co-location network of participants, which is inferred by using Bluetooth scan hits. For each pair of users, we counted the number of co-location scans, and used this as a proxy for the actual time that they spend

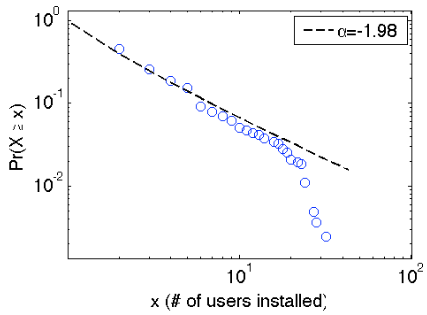

Figure 2 - Distribution of the number of users per app.

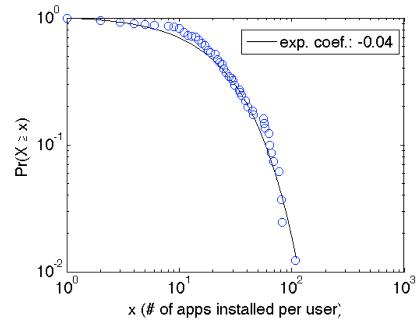

Figure 3: Distribution of the number of apps per user.

in a physical proximity to each other. We removed the recorded Bluetooth hits between midnight and 7am every day, since devices in neighboring apartment might sense one another, which may be incorrectly recognized as social interactions. We generally saw that husband and wife have over 1000 co-located BT scan events after the removal. Besides Bluetooth scan hits, for each pair of participants we also counted the number of common apps installed on both phones.

We divided the dataset into two groups by the threshold of 10 Bluetooth hits, which is at most 1 hour of co-location for the month used in this project. Group 1 with Bluetooth hits ranging from 0 to 10, which we assume to be mostly strangers and distant acquaintances. Group 2 are pairs with Bluetooth hits ranging from 10 to 2000. We then conducted a 1-way ANOVA and K-S test on the distribution of shared apps for every pair in both groups. The result is presented in Table 1(a). The mean of the shared apps for pairs in Group 1 is much less than the mean of Group 2. Both tests strongly reject the null hypothesis that the numbers of common apps are under the same distribution for both groups.

The second network we investigated is the self-report network. In the beginning of the study, each participant was asked in a survey to label other participants on a closeness scale of 0-10. We then created an adjacency matrix based on all self reports, and calculated the common apps shared by every pair of participants. For each pair, the closeness measure in this result is defined as the average rating from the two participants on each other. We also divided pairs into two groups: Group 1 includes pairs with closeness measure less than 1, and Group 2 is in the range (1, 10]. Therefore, Group 1 consists of strangers together with distant acquaintances, and closer relationships will all

(a)	Group 1	Group2
BT Co-Location Closeness Range	[0,10]	(10,2000]
Mean #Common Apps / Pair	2.7253	4.9
ANOVA: F=74.48, p<0.0000001		
K-S test: True, p=7.8e-19		

(b)	Group 1	Group2
Self Reported Closeness Range	[0,1]	(1,10]
Mean #Common Apps / Pair	4.75	4.05
ANOVA: F=4.97, p<0.026		
K-S test: True, p=0.0045		

Table 1: Summary results for (a) Bluetooth proximity closeness and (b) Self reported closeness.

be included in Group 2. As with the Bluetooth collocation networks, means, ANOVA, and K-S test results are shown in Table 1 (b). ANOVA shows borderline significant difference in the numbers of common apps from both groups, but less strong than the BT proximity network. K-S also shows similar results. However, the mean number of common apps is 4.97 for group 1 and 4.05 for group 2, suggesting that the two groups share almost same number of apps, with the stranger group sharing even more common apps. We then tested the border threshold with other values between [0, 2], and notice little difference in the means and the two statistical tests.

Discussion

In conclusion, we discovered that people who spend more time in face-to-face interaction are more likely to share common apps. In fact, in our dataset, pairs with face-to-face interaction share on average two more common apps on their phones compared with pairs with little face-to-face interactions. Those face-to-face interactions might include group activities, religion-related interactions, time spent with significant others and many other possibilities. However, we also observed that the self-reported friendships do not result in an increase in the number of common shared apps. We believe our results provide strong evidence on app diffusion patterns: apps do spread via social interaction. In particular, the diffusion of apps relies much more on the face-to-face interaction ties than the self-perceived friendship ties. Therefore, one should be cautious in using declared friendship networks to infer the spreading of smart-phone apps and for applying viral marketing strategies, since the face-to-face interaction seems to have a stronger correlation with app diffusion.

References

[1]N. Eagle, et al., "Inferring Social Network Structure using Mobile Phone Data," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 106, no. 36, 2009, pp. 5274-15278.

[2]D. Lazer, et al., "Computational Social Science," *Science*, vol. 323, no. 5915, 2009, pp. 721-723.

[3]I. Chronis, A. Madan, A. Pentland, "SocialCircuits: the Art of Using Mobile Phones for Modeling Personal Interactions," 2009, pp. 1--4.

[4]O. John and S. Srivastava, "The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives, " 1999.

[5] P. Dumpala, "The App Market Will Be As Big As The Internet in 2020, " The Business Insider, 2009.

[6] http://www.androlib.com, July, 2010