Analysis of Mobile Phone Call Data to Determine User Characteristics and to Enhance User Experience

S.R. Subramanya
School of Engineering, Technology, and Media
National University
San Diego, CA

ABSTRACT
Mobile phones have become ubiquitous in a very short time, having the highest penetration among all known technologies, with over 5 Billion users worldwide, and growing. Even the feature phones generate a variety of data during normal usage of the phone, while the smartphones generate even greater amounts of data with high-speed data accesses. Analyses of these data in combination with rule bases and heuristics would yield a wealth of information on the characteristics, personality, and lifestyle of the user. In this paper, we present a scheme which uses analysis of a typical mobile feature phone call data to yield interesting and useful results which would enable determination of some of the mobile user characteristics and enable possible productivity improvements.

General Terms
Mobile phones; Call data analysis.

Keywords
Mobile phones; phone call data; mobile user characteristics; mobile user experience.

1. INTRODUCTION
Mobile phones have witnessed the most rapid growths and penetrations. With over five billon current mobile phone users worldwide and growing, they have become ubiquitous and powerful. The more recent and widely popular smartphones are rivaling the PCs and laptops of not too long ago in terms of their capabilities. For example, they have processors with clock speeds ranging from 620 MHz to dual core processors of 1.5 GHz, 128 MB to 512 MB of RAM, 4GB to 64 GB of internal flash memory, display resolutions of 320x480 to 960x640 pixels, and 2 – 8 Megapixel cameras with 1080p video recording.

In this paper, however, we consider the feature phones (which still comprise about 70 – 75 % of market). Weighing about 3 – 4 oz. and conveniently fitting in one's palm and in shirt pocket, today's mobile feature phones (as opposed to smart phones) have processor cores with speeds of about 125 MHz, internal RAM of about 64 – 128 MB, a flash (non-volatile) memory of about 128MB – 512 MB, and external memory (Mini-SD) cards of about 512 MB – 2 GB. Most have a main LCD display of 128x160 to 240x320 pixels, with 65K colors. The battery usually provides about 200 hours of standby time and about 4 hours of talk time.

Other than their primary function of phone calls, they have increasingly become multifunctional. They support messaging (SMS, MMS); capture still pictures and short video clips using built-in cameras (with resolutions of about 1.3 – 2.0 mega pixels); music players which play music files, stored in internal flash memory or external SD-card; voice recording; some built-in games. Native applications such as, calendar, schedule manager, alarm clock, notepad, calculator, games, voice memo, call counter, user selectable ring tones, ring-back tones, and wall paper are fairly standard features (see Fig. 1). Voice-activated dialing and voice memo facilities are also becoming commonplace. Most current models are Bluetooth enabled, supporting wireless headsets and wireless data transfer to PCs.

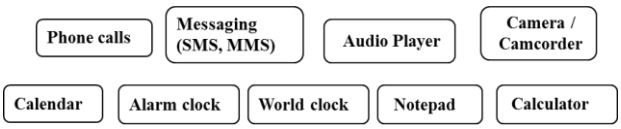

Fig 1: A few of the major functionalities and applications in a typical mobile feature phone

Given so many different functionalities, even routine uses of the mobile phone generate fair amounts of diverse data which are stored in a file system in the phone's non-volatile (flash) memory. In this paper, we consider only the call log data and analyze them to derive numerous pieces of information, which, in conjunction with some heuristics and a rule base, could be used to derive interesting information which could be used to make educated guesses about the user's nature, lifestyle, etc. Analysis of the various data generated and stored in mobile phones could lead to discovery of interesting, sometimes unexpected, information/knowledge. These could be used by other 'higher-level' applications in social networking, lifestyle monitoring, and the like. These pieces of information together with their visualization could be used effectively to provide enhanced user experience as well as lifestyle and productivity improvements.

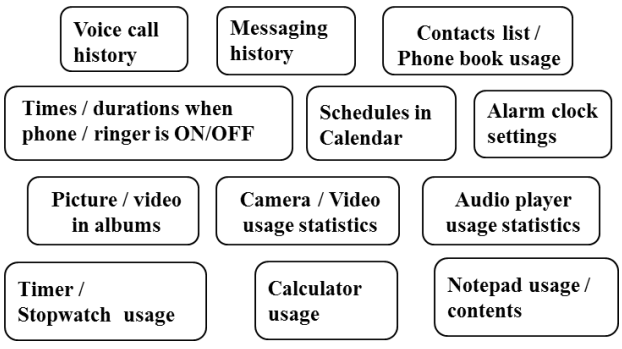

Fig 2: A few of the various kinds of common data that are generated in a typical mobile feature phone

There are a variety of data generated from the use of different applications in a mobile phone. A few of them are shown in Fig. 2. For example, using the mobile phone for voice calls generates call history which contains the phone numbers of the incoming, dialed, and missed calls. It also contains the times and durations of the dialed and incoming calls. Similarly, use of the mobile phone for other applications generates a variety of data related to the use of the application.

Simple (elementary) data analysis schemes applied to data generated from the use of some mobile phone application may yield statistically interesting facts.

For the purposes of this paper, we consider only the call log data. As part of the future work, we plan on the application of sophisticated analyses, together with the use of heuristics and rule bases, to combinations of data from more than one application to derive more interesting information and patterns of usage and behavior, which would not be possible using any of the individual data sets alone. These could be of tremendous help in time and lifestyle management and improvement.

The next section briefly describes a few samples of related work. Section 3 describes the proposed call data analysis schemes. Section 4 presents the results of analysis of the experimental data. Several planned extensions and future work is given in Section 5, followed by conclusions.

2. RELATED WORK

There has been some work done recently in the analysis of mobile phone data for a variety of purposes. Novel aspects of human dynamics and social interactions are investigated using analysis of extensive mobile phone data with respect to both time and space in [1]. The data in mobile devices found at crime sites might provide investigators several valuable pieces of information. The field of digital forensics which analyzes any digital data at the crime scene or related to the crime now encompasses mobile phone data as well. A survey of numerous mobile phone technologies, hardware, and analytics tools is given in [2]. has grown out of the mainstream practice of computer forensics. Practitioners are faced with various types of cellular phone generation technologies, proprietary embedded firmware systems, along with a staggering amount of unique cable connectors for different models of phones within the same manufacturer brand. An analysis of mobile phone usage data of the county provided phones to its employees for the purposes of tracking usage and accountability by the Montgomery county, Maryland is given in [3]. Using Bayesian location inference framework upon massive mobile phone data, a methodology of detecting large events (ex. rock concerts, sports finals, etc.) has been described in [4]. An analysis of location information of mobile phone users gathered over two months, at discrete times when the phones are active, in Los Angeles and New York has been performed, and significant differences in the mobility patterns have been discovered in [5]. Digital information that are implicitly generated in mobile phones have been analyzed to derive associations between locations and travel demands for different types of social events in [6].

3. ELEMENTARY CALL LOG DATA ANALYSIS

The elementary data analysis described here gathers simple statistics on a single data set – the phone call data – in mobile phones. For example, consider the voice call history consisting of elementary data such as (a) phone numbers of incoming, dialed, and missed calls, and (b) the days, times, and durations of incoming and dialed calls. Using elementary data analysis on the above data, one can derive various distributions related to calls as shown in Figure 3.

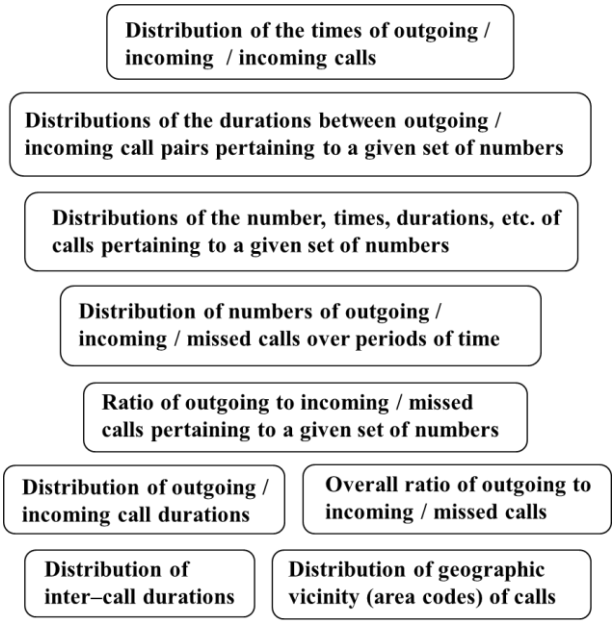

Fig 3: A few of the statistics that could be gathered from data related to calls

Simple inference rules applied to the results of the elementary analysis as shown in the outline in Figure 4 could yield interesting and useful information – some of which could already be known, while several other could be surprising discoveries. Some interesting facts that could be derived (inferred) from the above results of analysis are (a) the most/least called numbers, (b) the persons who call this phone most, (c) the numbers with the longest/shortest talk durations, (d) geographic distributions (area codes) of the numbers, (e) promptness of the person in returning missed calls, (f) importance of the calls, etc.

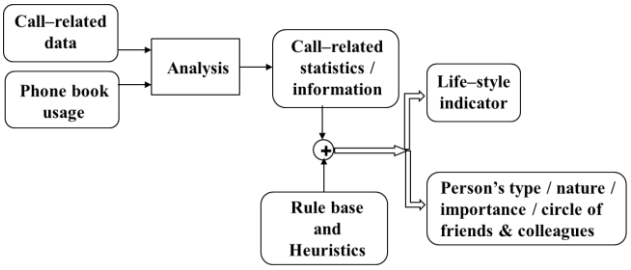

Fig 4: Visualization of distribution of call durations and inter-call times

The information about the most called/received numbers could be used advantageously to automatically provide a more personalized and enhanced communication experience. For example, if the call from a 'favorite' or 'important' person cannot be accepted by the called party (due to called party being busy, phone being turned off, etc), an automated message (SMS or personalized voice message clip) could be sent by the system to the calling party to update the status of the called party. Another example would be to display pictures, animations or video clips related to the (intimate) person which are context-appropriate. These are pre-stored in the phone and thus do not use precious bandwidth.

Further analysis of the data on the number of handovers and number of roaming situations (when the handset is in a network other than its home network) would give an indication of the mobility of the user. A metric called 'mobility index' might be derived which could be used to

monitor or even optimize mobility situations. Detailed studies of aggregates of the call patterns could be used for the provisioning and allocation of communication resources dynamically in a cost-effective manner.

4. EXPERIMENTAL DATA AND ANALYSIS

Due to the highly personal and private nature of mobile phone data, it is very difficult to get people to share their data. A well-established mechanism for anonymizing the sensitive data is required in order to gather and use mobile phone data from across user groups. In this study, data from mobile phone call logs for a typical home maker in the US was analyzed for some salient overall characteristics such as (a) total number of calls (outgoing, incoming), (b) call durations, (c) inter–call intervals, (d) ratio of outgoing to incoming calls, etc. over a period of 6 months. Not much variation was found in these data across months. The call log data for the results presented in this paper were taken over a 4 week period. Some of the analyses of the call data are given in the following sections.

4.1 Aggregate of phone calls over a 4-week period

Figure 5 shows the distribution of calls and durations over a 4 week period.

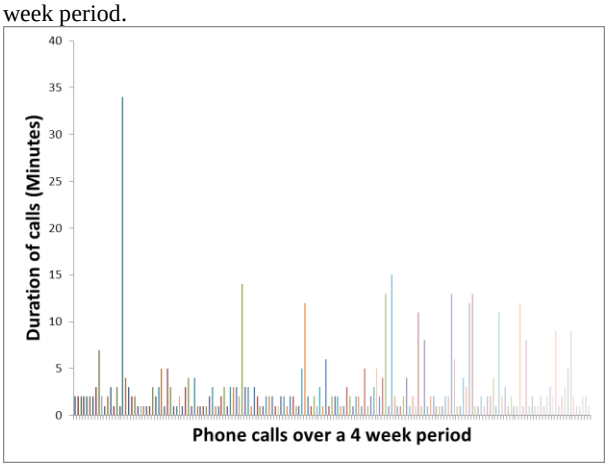

Fig 5: Duration of each call plotted over a 4-week period

Observations: The distributions of phone call durations are bursty – the calls are mostly of short durations, and once in a while there are high peaks corresponding to calls of long durations.

4.2 Call durations across phone numbers

Figure 6 shows the plot of total duration of calls (both dialed and received) for each of the phone numbers in the call log over the 4-week period.

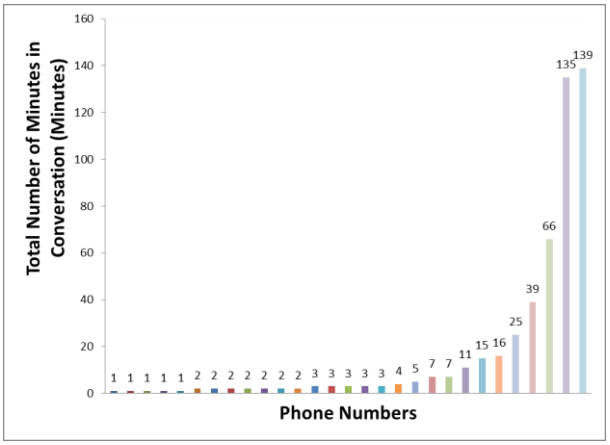

Fig 6: Total duration of calls V/s phone numbers

Observations: There is a very small percentage of phone calls with long durations – corresponding to known persons with close relationship and relatively long tail of short call durations.

4.3 Distribution of call durations over weekdays and weekends

The plot in Figure 7 shows the distribution of the aggregate of the call durations over weekdays and weekends in the 4-week period.

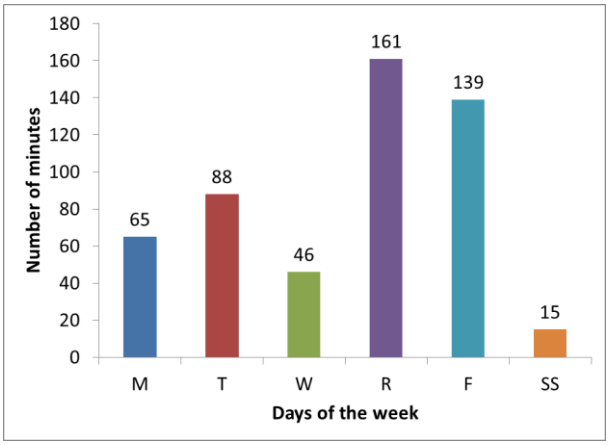

Fig 7: Durations of calls on weekdays and weekends

Observations: The week starts off slowly with respect to the total duration of calls on Monday. It has high peaks on Thursday and Friday. The time spent on weekends is very minimal. Time spent talking on Wednesday is the lowest among weekdays – a very likely reason being the user involved in part-time work or some serious activity on that day.

4.4 Distribution of call durations across all calls

The plot in Figure 8 shows the distribution of the call durations across all calls in the 4-week period.

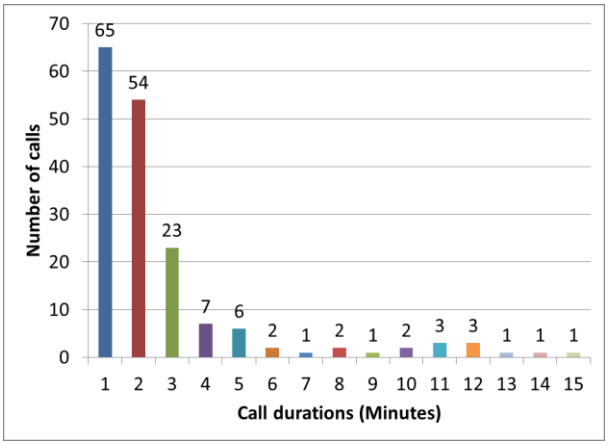

Fig 8: Number of calls v/s call durations

Observations: The number of calls of short durations are very high (for 1 – 2 minutes) and tapers off rapidly for durations of 3 minutes and above.

4.5 Distribution of call durations at different times of the day

The plot in Figure 9 shows the distribution of the aggregates of call durations over 1-hour intervals over the 4-week period.

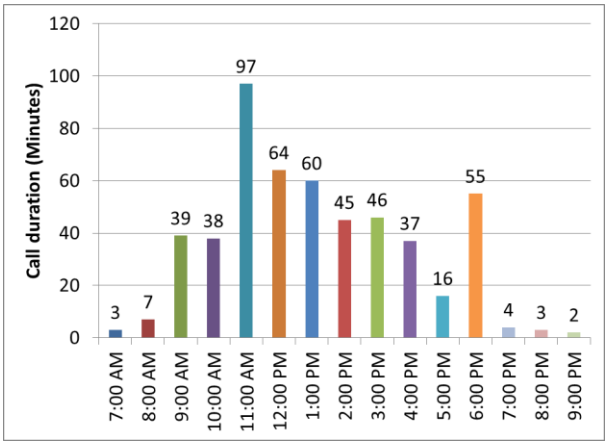

Fig 9: Call durations at different times of the day

Observations: The call activity starts about 7 am. It ramps up and peaks between 11:00 am and noon. Generally there is high activity between 9 am and 5 pm, and between 6pm and 7 pm. Between 5pm and 6 pm there is lesser activity.

4.6 Geographical distribution of number of calls and call durations

The plot in Figure 10 shows the distribution of the geographic locations of the phones (actually, in this case, this is the location where the phone is registered; this is reasonable given the fact that almost all users in the log would not roam too far).

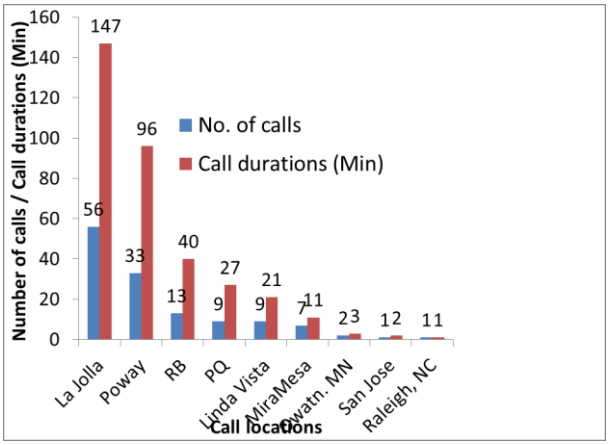

Fig 10: Geographical distribution of number of calls and call durations

Observations: Most of the calls and correspondingly the longer call durations are within a small geographic vicinity of 4 – 5 locations.

4.7 Distribution of inter–call durations

The plot in Figure shows the distribution of the inter–call durations over the 4-week period. The duration between calls across day boundaries are not considered (ex. the duration between the last call of a day and the first call of the next day).

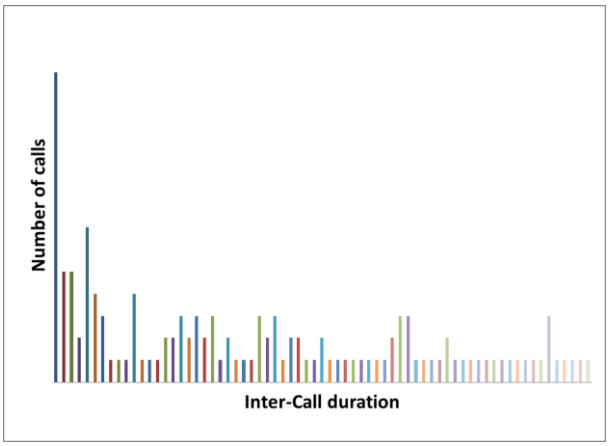

Fig 11: Distribution of inter-call durations

Observations: The distribution of inter–call durations follows an exponential decay with a long tail. There are a large number of calls with very small inter–call durations. The long tail consists in the aggregate of calls with longer inter–call durations being a large number.

5. FUTURE WORK

The planned future work consists of analyses of the following: (a) call data for different personas; (b) in-network and roaming calls; (c) messaging (SMS) data; (d) GPS/location data; (e) camera/camcorder usage data; (f) Web usage data; (g) calendar / schedule data, and (h) development of rule bases and heuristics to use combinations of the above data and their analyses to derive more interesting characteristics and behaviors of individual users and also group behavior.

5.1 Extended data analysis

Application of sophisticated analyses to combinations of data from multiple applications, together with the use of some heuristics and rule bases, would yield interesting and useful information and knowledge. These could be beneficial in monitoring, managing, and improving lifestyles. A few examples are given below with just brief outlines due to space considerations. Using rule base and heuristics on suitable (a) statistical data about calls and messaging in conjunction with the phone book (list of contacts), (b) audio player and camera usage statistics, and (c) the use and contents of calendar, information about the person's nature could be derived. For example, if the ratio of outgoing to incoming calls is high, the times of calls are uniformly spread out, calls are made to a good proportion of the persons in the contacts list, the camera is used frequently along with MMS messaging, etc, then, it is very likely that the person is socially outgoing. If the ratio of incoming to outgoing calls is consistently high, if the number of missed calls is significant, if the number of voice mails is considerable, if the calendar has several meetings/appointments, then it is likely that the person has a relatively higher level of responsibility.

Figure 12 shows some of the information which could be used to derive the list of favorite or important persons relative to a given person. For example, if the numbers of the dialed and received calls are evenly matched, the times of the dialed calls of a number are not too far off from the times of the corresponding missed calls, the calls have the similar regularity of occurrence even under roaming, etc, then these are indicative of a favorite person. Figure 7 shows some of the data which could be used to determine a person's work-life balance. For example (assuming this is a working adult), if the same set of (work) numbers are called during weekdays as well as weekends, if the calendar has a fair number of meetings on a weekend day, considerable number of messaging on weekend day(s) to/from the same numbers used during week days, then it is quite likely that the person is involved with work during weekend day(s) too.

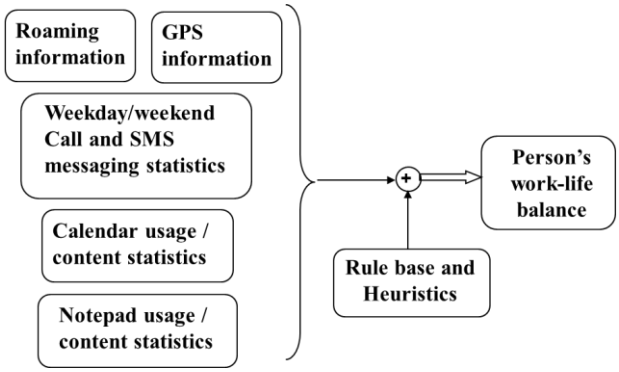

Fig 12: Total duration of calls V/s phone numbers

5.2 Visualization of analyzed data

Visualization of huge amounts of information has the benefits of providing the humans with some patterns and insights which tables of numbers or elementary graphs cannot provide. An outline of the visualization scheme is shown in Figure 13.

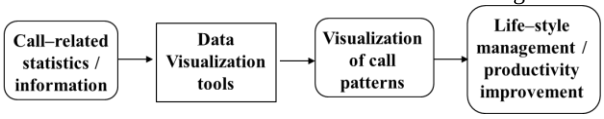

Fig 13: Outline of visualization of call-related information

Higher–level information is obtained by extended analysis of basic data, and visualization is applied on this higher–level information.

The visualization of a small sample of the distribution of the call durations and the times between calls using a simple tool is shown in Figure 14. The visualizations could be of data aggregated over a sufficient period of time. Such visualizations would easily reveal some patterns of calls. For example they indicate the levels of bursty-ness of calls, whether the ratio of incoming to outgoing calls is high or moderate, etc. Note that if the times between the calls are uniformly distributed with small inter-call idle times then it would be counterproductive (assuming that phone calls are not the primary productive activity). Therefore, some of the seemingly unnecessary calls could be either deferred or missed. Thus simple data analysis together with visualizations could be used beneficially to monitor, modify, or improve several lifestyle activities.

Fig 14: Visualization of distribution of call durations and inter-call times of different kinds of calls

6. CONCLUSIONS

Routine use mobile phones generate a variety of data, which by themselves may seem uninteresting. However, using both elementary and extended analyses on combinations of the data yields interesting information. Together with rule bases and heuristics, further information and knowledge could be derived which could be used in higher-level applications in social networking, lifestyle monitoring, and even resource optimization.

The mobile phones of the future are expected to play key roles as assistants, companions, tutors, caretakers, and lifesavers, in addition to being devices for communication and entertainment. Some of the ideas developed in this article could be used to analyze rich sets of data/information which may be used and extended by human interaction designers, sociologists, engineers, and computer scientists to design schemes for enhancing the overall experience of mobile phone users.

7. REFERENCES

[1] J. Candia, et.al. 2008. Uncovering individual and collective human dynamics from mobile phone records. J. Phys. A: Math. Theor. 41. 11pp.

[2] S. G. Punja and R. P. Mislan. 2008. Mobile Device Analysis. Small Scale Digital Device Forensics Journal, 2(1), June 2008 ISSN# 1941-6164

[3] http://www.montgomerycountymd.gov/content/EXEC/stat/pdfs/1-19-10_ppt.pdf

[4] http://hal-ens.archives-ouvertes.fr/docs/00/62/71/22/PDF/ socialcom_2011.pdf

[5] S. Isaacman, et. al. 2010. A Tale of Two Cities. Workshop on the Analysis of Mobile Phone Networks MIT, Cambridge, May 2010, 50-52.

[6] F. Calabrese, et.al. 2010. Analyzing cell-phone mobility and social events. Workshop on the Analysis of Mobile Phone Networks MIT, Cambridge, May 2010, 59-61.