Understanding Call Logs of Smartphone Users for Making Future Calls

Mehwish Nasim[1], **Aimal Rextin**[2], **Numair Khan**[3], and **Muhammad Muddassir Malik**[3]

[1]University of Konstanz, Germany, mehwish.nasim@uni-konstanz.de

[2]COMSATS Institute of Information Technology, Islamabad, Pakistan, aimal.rextin@comsats.edu.pk

[3]National University of Sciences & Technology, Islamabad, Pakistan, {numair.khan, muddassir.malik}@seecs.edu.pk

ABSTRACT

In this measurement study, we analyze whether mobile phone users exhibit temporal regularity in their mobile communication. To this end, we collected a mobile phone usage dataset from a developing country – Pakistan. The data consists of 783 users and 229,450 communication events. We found a number of interesting patterns both at the aggregate level and at dyadic level in the data. Some interesting results include: the number of calls to different alters consistently follow the rank-size rule; a communication event between an ego-alter(user-contact) pair greatly increases the chances of another communication event; certain ego-alter pairs tend to communicate more over weekends; ego-alter pairs exhibit autocorrelation in various time quantum. Identifying such idiosyncrasies in the ego-alter communication can help improve the calling experience of smartphone users by automatically (smartly) sorting the call log without any manual intervention.

Author Keywords

Mobile Phones; Call Logs; Time Series; Temporal Patterns

ACM Classification Keywords

I.5.2 Design Methodology: Pattern analysis; H.5.2User-centered design

INTRODUCTION AND MOTIVATION

In this work we focus on mobile calling as an example of demonstrating temporal patterns in communication and also discuss possible benefits of discovering peculiarities in this domain. An estimated 261 million Americans own mobile phones, with a daily average of almost 1.3 billion mobile communication events [24]. Hence mobile phones represent an important communication medium. A call can be made in a variety of situations because of mobile phone's portable nature and one can assume very little about the context of a call. These two factors, frequency and versatility of use, necessitate an extremely efficient call-making interface design. Users generally make phone calls in two ways: either by selecting the callee from a contact list, or through the call log. The former displays contacts in alphabetical order with no consideration of past calling behavior. While most mobile phones offer the capability of selecting certain contacts as *favorites*, the favorites list is, however, still a static list, requiring active intervention by the user in order to update. Call logs, on the other hand, do take past user behavior into account, displaying called numbers in reverse chronological order. The model of user behavior assumed by call logs is, nonetheless, highly simplistic. It supposes that the likelihood of calling a particular contact, $P(c)$, is a monotonically decreasing function of the time elapsed since last contact. Sociologists have, however, shown that human life is temporally organized and that most social interactions have fairly reliable temporal regularity [45]. This implies that $P(c)$ could be estimated to a certain extent by understanding user calling patterns. Such an implication, correct, would allow for the design of a considerably more efficient calling interface than what is provided by either contact lists, or chronological call logs.

We began this work by conducting a small scale pretest. We first collected online survey-responses from 28 participants who own Android phones and then analyzed call data from 13 of those participants (Data statistics are available in Appendix A). The survey aimed at getting a deeper insight into the calling behavior of users with respect to the use of call logs for making future calls. Survey consisted of questions such as: whether the participants call different people on different days of the week; whether he/she calls different people on weekends; whether they use missed calls to indicate any kind of signal; how often they use call log to dial a number, etc. Once the participants finished the survey, they were given the option to share their data for research purpose. Only 13 participants agreed to share the data. We then sent a link to our app hosted on Google Play to those participants. Once the app was installed, it automatically submitted an anonymized version of the data to our server. Most survey participants agreed that they usually use call logs to make future calls. Except for four participants, everyone else had experience in using missed call as a signal, for instance as an indication for the other person to callback. More than 50% of the participants agreed that they call different people on weekends as opposed to weekdays. Analysis of their call log data indicated

MobileHCI '16, September 06-09, 2016, Florence, Italy

DOI: http://dx.doi.org/10.1145/2935334.2935350

that 30% ego-alter [1] pairs communicated more on weekends as compared to the weekdays.

Encouraged by the qualitative as well as the quantitative results of the pretest, we collected a dataset from the general population to get further insight into the calling behavior. We were interested in a number of research questions related to understanding the characteristics of communication behavior on mobile phones, and patterns of communication at individual level as well as at a finer level i.e. ego-alter interaction. We were interested in knowing: how many active contacts do mobile phone users have? how often they are called? With respect to historic logs, we were interested in finding: Distribution of calls, more specifically, what percentage of communication goes to top contacts? and how often people call the recently called contacts?

Contributions:

1. Collection of an original mobile phone usage dataset of 783 users with 229,450 communication events. The dataset captures call data from an understudied population: Pakistani mobile phone users. This is a one of the potential strengths of our work[2].

2. Analysis of the collected data to answer the research questions posed in the last section. We start at an aggregate level, studying aggregate patterns and discussing our findings. We then move on to the dyadic level, and focus on the variation that is hidden at the aggregate level, i.e. individual differences and marked daily activity patterns.

RELATED WORK

Temporal regularity can be observed in time variation of activity on online social networks such as Youtube, Twitter and Slashdot, and also in frequency of edits made on Wikipedia [14, 20, 44]. Activity on twitter in various languages shows that circadian patterns exists for tweets all around the world [38]. Temporal interactions have been used to study human behavior for instance commenting behavior of Facebook users (a consequence of social selection or social influence effects)[28]. Temporal interactions have also been used to predict links in social networks [23, 27, 37].

Call log data has been shown to hold significant potential of providing insights into the underlying relational dynamics of societies, evolution of relationships over time and, in the absence of survey data, the quantification and prediction of social network structures [13]. Data of calling patterns has been used to infer friendships relations and uncover individual and collective human dynamics [13, 10, 18, 22, 12, 29]. Call-volume data has been used to explore whether the distribution of calls in an urban population follow routine patterns or not, and whether the variation of such patterns in different parts of the city can be explained [36]. Inspired by effective studies on calling patterns, researchers have devised several call prediction models. In [33], authors predicted the outgoing and incoming calls on Reality Mining dataset [12] based on most recent calling data. Out of the 94 datasets, they used a small subset of 30 users for performance evaluation. Barzaiq et al. [6] modeled the historic call patterns of users and achieved a 35% accuracy for call prediction on synthetic data. Haddad et al. [15] discuss a probabilistic model that uses call frequency to predict incoming and outgoing calls for each individual contact. Recent studies of human behavior indicate that the timing of communication events is characterized by long dormant periods interspersed with bursts of high activity [5, 19, 43]. Barabasi [5] attributes this bursty non-Poisson character of human behavior to a priority-based queuing process. This view is supported by Jo et al. [19] who show that burstiness remains in mobile communication data even after circadian and weekly patterns have been removed, precluding the attribution of periods of inactivity to nights or weekends. They conclude that burstiness results from non-homogeneity in human task execution mechanisms. Kim et al. [21] conducted a study on a large dataset from North-American mobile phone users. The results suggest that the caller-callee behavior cannot solely be modeled using the Poisson distribution. Based on frequency of information exchange between the users, they classified the user-pairs into three categories characterized by the inter-arrival times between calls made between pairs. In a related study, Cardillo et al. [11] studied human proximity patterns in two data sets: the Reality Mining dataset and the co-location traces from INFOCOM'06. They found that proximity patterns from the MIT data contain both weekly and daily periodicity; most probably a result of how academic activities are scheduled at a university, while the INFOCOM'06 data showed only daily periodicity. Caridillo et al. extended this observation to study how cooperation emerges in a human society.

A patent from Google suggests that an adaptive contact list may detect contextual information for a given mobile phone user and may identify appropriate contact entries [17]. While studying the effects of two different UI adaptation techniques on user performance, Tsandilas and Schraefel [39] conclude that adaptation is always more effective, even when the accuracy of prediction is low. Bentley and Chen [7] found that the majority of contacts in a modern aggregated mobile phone book are rarely used. Their study shows that the five most frequently contacted alters represent 80% of phone and text communication. Bentley and Chen suggest a redesign of the content and representation of contact lists. A redesign of contacts book was proposed also in [30]. The data for Bentley and Chen's study was collected from user in the United States via an Android app. Volunteer bias especially as a survey was also required from the users. Moreover, while representative of the general population of the US, the authors acknowledge that communication patterns in other parts of the world may vary.

While studying social network turnover, Aledavood et al. [4, 3] found that individual calling and messaging behavior follows a circadian rhythm. Their study of 24 subjects revealed that the frequency and entropy of communication displays a distinct daily pattern that remains persistent over time. Findings on temporal patterns in Aledavood et al. [4, 3, 2] are

[1] Ego is the focal actor[16] who has installed the app. Alters are people with whom ego communicates using voice calls.

[2] Anonymized data can be obtained by emailing at aimal.rextin@comsats.edu.pk.

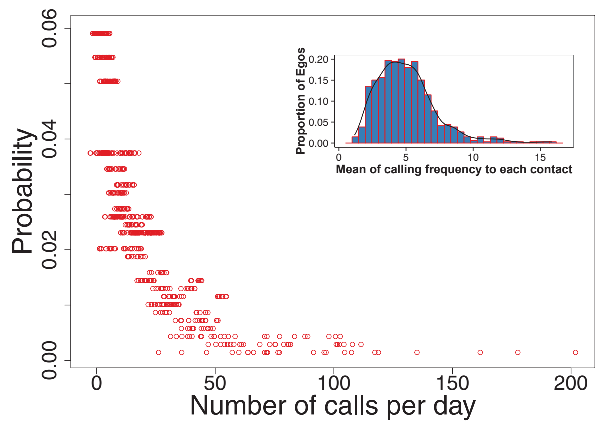


Figure 1. Average number of calls per day and the overlay plot at the top right shows mean of average number of calls per user per contact.

attributed to the diurnal cycle of human beings. Moreover, it was found that frequently called contacts are the ones most likely to be contacted during low entropy periods. Nonetheless, the studies did not answer the question whether communication between pairs of individuals is periodic.

METHOD

We developed a smartphone app specifically for the purpose of this experiment. Data collection was limited demographically to users of smartphones running the Android OS, and geographically to the country of Pakistan.

The data was collected by marketing the application through Facebook ads, word of mouth and wall posts on technology pages on Facebook. Pakistan ranks 8^{th} in the list of number of mobile phones in use in the country, ahead of Japan, South Korea and all European countries[42]. In December 2015, the number of 3G/4G subscribers exceeded 23.16 Million [34]. While industry sources estimate that Android users represent 68% of the total smartphone population in that country, extensive market surveys are lacking and, hence, conclusive judgments about the qualitative nature of the sample cannot be made.

Pakistan is a low income country and people are interested in reducing their mobile usage cost [1]. To make its value proposition more attractive, the application presented users with the most economical telecom service for their needs based on past calling behavior. These telecom services - also referred to as "packages" in the local parlance - differ primarily in the calling rates they offer during specific hours of the week. A recommender system for similar telecom products was developed by Zhang *et al.* [46]. But, where they used fuzzy-set techniques to select the most economical product, our recommendations are based on a simple simulation run with the users' call history. Including this additional functionality in our data-gathering app not only expanded our sample set, but we also expected it to mitigate the volunteer bias natural in such survey data collection methods. Users were notified that their call data would be used for academic research purposes.[3] In case users did not agree with the privacy policy, they could quit the app without compromising their data. Since the data was collected on mass scale, collecting demographic information at that scale was not feasible. We call this data as the *Paki-Smartphone dataset*.

AGGREGATE DATA ANALYSIS

We uploaded our application on Google Play on July 28, 2015. The process of data collection lasted from July 28, 2015 till September 24, 2015. The application did not replace any functionality on the host phone and did not interfere with normal usage of phone in any way. The application collected the historic data from call logs of smartphone users. The event data consisted of the following fields (anonymized): unique id of the phone, contact number of the alter, communication event type, i.e. received call, missed call, outgoing call, etc., date and time of the event and duration of call. The data collected by the application had timestamps from April 19, 2015 till September 23, 2015. The data consists of 783 users with $229,450$ communication events and more than $12,000$ active contacts.

Of the total calls, 24% calls were incoming and answered, 19% were incoming and missed calls, and 54% of calls were outgoing calls. These call statistics are very interesting, firstly, because Bentley and Chen [7] also reported similar statistics for their dataset of 200 users, and secondly, they are comparable with the statistics of the 13 users from our pretest. We also looked at the statistics from the Reality Mining dataset that was collected at MIT. There were 22% incoming calls and 66% outgoing calls. However, the percentage of missed calls was only 10%. We also found that 2% calls were incoming and rejected calls in our dataset. None of the calls were voice mail calls, whereas, a minuscule number of calls were from the refused list. The average length of incoming phone calls was around 104 seconds with a median value of 38 seconds. Average length of outgoing calls was 56 seconds. A handful of calls were very long. The longest call lasted one hour. Only 1.6% calls lasted more than 10 minutes. About 83% of the communication took place between 6:00 a.m. and 9:00 p.m. It was especially interesting to note the relatively high percentage of missed calls. Although, a high number of missed calls was also reported in [7] but the authors could not mention a plausible reason behind it. This statistic for missed calls is a noticeable artifact of our dataset from a low-income country, where users, sometimes give missed calls to indicate a signal - an easy way to save money. One of the participants in our user study stated, "*Whenever I reach home late, I give a missed call to my mother so that she could open the door*". Another participant indicated that he regularly talks to his girlfriend in the evening. *"When her parents are around, she gives me a missed call which means that I am not to supposed to call her"*(a very typical setting in a South-Asian society). The notion of missed calls is in the core of Pakistani

[3]Privacy Policy: Package Advisor collects information related to dialed and received calls in order to propose better call packages. We may use this data later for improving our application as well as for academic research purposes. However, we will not share any collected data containing personally identifiable information with third parties.

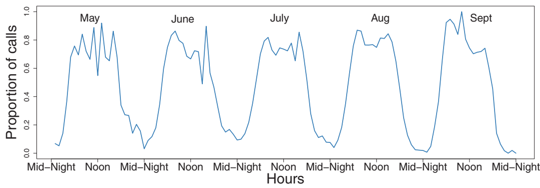


Figure 2. Number of events per hour for each month in the Dataset at an aggregate level.

mobile phone market to such an extent that Unilever launched a campaign called "missed-calls" on mother's day. The campaign attracted over 1 Million calls in 2015 [26].

Distribution of calls

Bentley and Chen [7] observed that most calls are to/from 5-10 contacts.This view is also supported by [21]. Similarly, Bergman *et al.* [8] observed that the participants of their study did not call 47% of their contacts for 6 months. We empirically found a similar but more interesting pattern; every user's call distribution very closely follow the equation below:

$$\frac{e^a}{x^b} \tag{1}$$

Here, a and b are real number that is fixed for each participant and x is the rank of the alter that varies from 1 for the alter with the most communication events and so on till the rank of the alter with the least communication events. It is worth noting that a and b both lie in a narrow range as a varied between 0-7 and b varied between 0-2.5. We observed that our equation fits the the data very well and we got a mean adjusted R^2 of 0.89 and the standard deviation was 0.16. Note we removed the data of 27 egos because their number of communication events was below 4. It is interesting to note that cities and their rank also follow a similar distribution and this pattern is generally known as the *rank-size rule* [35]. Equation 1 indicates that any future redesign of the contact list would probably need to compute the important contacts for each individual. The number of important(top) contacts can vary from about 5 for an individual with $a = 2$ to about 20 for an individual whose value of a is 6. We plan to further investigate why the Equation 1 varies from one individual to another and then apply that knowledge to improve calling experience of mobile users. The probability distribution function (PDF) of number of calls per user are shown in Figure 1. On average, each user made or received $\approx$ 22 calls per day. The mean of average number of calls per user per contact is plotted as a bar chart at top right in the same Figure.

Hourly and Weekly calling behavior

Our data consists of communication events between April and September 2015. Since, there were only eight events recorded in the month of April, we illustrated the aggregate number of events per hour for each month, from May till September in Figure 2. The communication activity is highest during the daytime hours and decreases by mid-night in every month. We also conducted the *Wilcoxon signed-rank test*, which is a

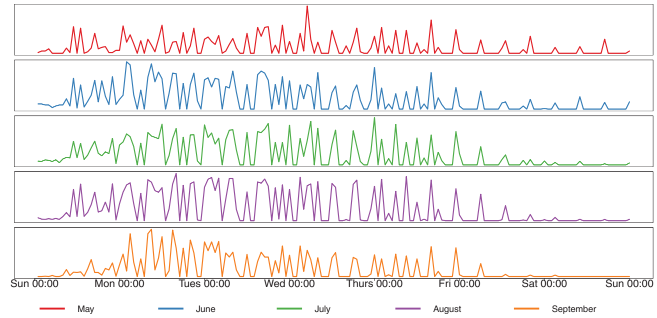


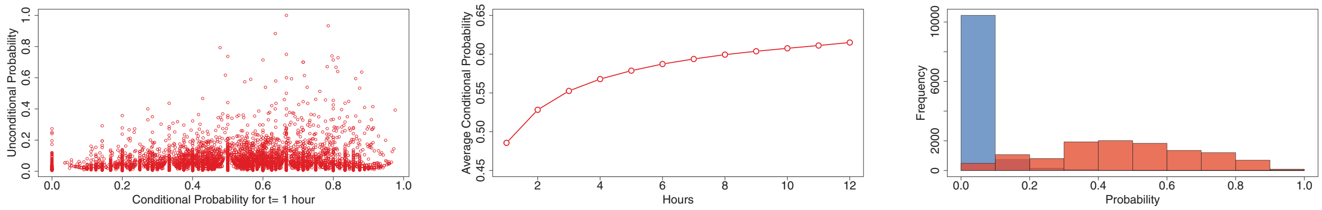
Figure 3. This Figure shows the calling patterns between Sunday and Saturday for each Month. The x-axis depicts the number of hours in each day for 7 days of the week. The difference between two ticks on x-axis is 24 hours (1 day). y-axis represents the aggregated proportion of calls made at each hour of the day by the egos. The distributions across all months look identical. Highest communication activity is observed between Monday and Thursday. Call volume decreases significantly, after Friday midday. Comparing all the distributions $\binom{n}{2}$ using Wilcoxon signed-rank test gave the following *p-values*: 0.96, 0.70, 0, 89, 0.41, 0.68, 0.85, 0.34, 0.78, 0, 20, 0.27**.**

non-parametric statistical hypothesis test where our null hypothesis was that the distribution of weekly calls for any pair of months is identical (Figure 3). The results show that the *p-values* are not statistically significant, hence insufficient evidence to reject the null hypothesis at $\alpha = 0.05$. This test is useful when the population cannot be assumed to be normally distributed. Only 26% calls were made on weekends in the entire dataset. A low activity during the weekend was consistent for all months. At a finer level we found that there was a higher probability of communication between 25% ego-alter pairs on weekends. Similarly, 75% ego-alter pairs were more likely to communicate during the weekdays in the dataset. This indicated that probably the nature of the ego-alter social relation is different for those people that a person calls on weekends versus those which he/she calls on weekdays. In the next section we further investigate the dyadic interaction patterns.

EGO-ALTER INTERACTION PATTERNS

Approaches on network analysis are exploring the conceptualization of networks as collection of data on dyads rather than on the graph. *"The distinctive characteristic of networks is that their units of observation (the identifiers of data points) are not single entities but pairs of entities, and that each entity may appear in multiple such pairs"*, Brandes et al. [9]. Studies on human communication behavior [2, 14, 38, 44], have investigated circadian patterns. They attributed the temporal regularity to diurnal cycle of human beings (when people are using modern communication modes such as mobiles and Internet) and that inter-individual differences arise due to geographical and cultural differences. We argue that the temporal communications patterns are not just a consequence of the diurnal cycle. At a finer level, we focus on the variation that is covert at the aggregate level, i.e. individual differences and marked daily activity patterns between pairs of users.

Probability of Calling an Alter Again



(a) Probability of a communication event between an ego-alter pair when there was another communication event between the same pair. For almost 97% it is clearly greater than unconditional probability of a communication event between that pair.

(b) The predictive power of conditional probability increases gradually and levels off very quickly.

(c) Distribution of unconditional probability (blue) and conditional probability (red) for all ego-alter pairs for $t = 1 hour$.

Figure 4. Weekly and Hourly Calling Activity in each Month.

Our informal user study showed that about 71% of respondents either always or usually use call log to initiate a call. If this trend is true in general, then we hypothesize that the probability of a communication event between an ego-alter pair should be significantly less than the conditional probability of a communication event between an ego-alter pair given that there was a communication event in the near past. Since we could not come up with a reasonable definition of near past, so we decided to check this hypothesis by computing the conditional probability between each ego-alter pair given that there was a communication event t hours ago, where $t \in \{1, 2, \cdots, 12\}$. Figure 4a clearly showed that a communication event is much more likely if there was a communication event within the last 60 minutes. In only 3% cases the unconditional probability was higher than the conditional probability. This result supports our hypothesis. However, this conditional probability does not increase significantly and levels off very quickly as we increase t as shown in Figure 4b. Moreover, Figure 4c shows a histograms of unconditional as well as conditional probability when $t = 1 hour$) for all ego-alter pairs. The distribution for conditional probability is normal and hence we estimate the mean and standard error ($L(\mu, \sigma)$) of this distribution using the Maximum Likelihood Estimate. We estimated $\mu = 0.48$ and $\sigma = 0.21$ as compared to the unconditional probability where $\mu = 0.041$ and $\sigma = 0.059$. To the best of our knowledge, we interpret these results as the first empirical evidence that call logs are fairly useful for making future calls. These descriptive statistics are also consistent with the survey results that we discussed in the Introduction.

Autocorrelation

Time domain periodicity detection methods make use of autocorrelation functions [31]. Autocorrelation refers to the statistical dependency between the values of a variable on related entities. In terms of time series data, like our dataset, autocorrelation implies persistence from one observation to another. Autocorrelation is a common characteristic of relational and social-network datasets; since mobile calling is also a form of social interaction, it is logical to test whether caller-callee interactions exhibit autocorrelation or not. In many time series, it is reasonable to expect that the m recent data points are likely to have an influence on the future data points. For comparison purpose, the sample for this experiment consisted of two datasets: the Paki-Smartphone dataset and the Reality Mining dataset of Eagle and Pentland [12] collected at the Massachusetts Institute of Technology. This latter dataset comprises call and text data for 94 egos. Each dataset was analysed independently. The data for each ego was grouped according to the contact the communication event was initiated to. We represented the individual communication events between the pair using a binary string, with each bit position representing a time quantum, modelled by a Bernoulli random variable Z. A bit (Z) was assigned a value of one if a communication event did occur during that time quantum, and zero otherwise. We have taken into account two types of time quantum —daily and hourly. Studies on human social behavior support our selection of time quantum. Human circadian rhythms intrinsically follow a period of approximately 24 hours [41]. Within this broad period, however, there are significant inter-individual differences which may correlate with distinctive temporal patterns of physiological and psychological variables, of gender and, personality traits [40].

We then determined temporal regularity for ego-alter pair using the Ljung-Box Q test ($\alpha = 0.05$) on the string. The Ljung-Box test, also known as a *portmanteau* test, is a function of the accumulated sample autocorrelations r_k, up to any specified time lag m [25] [4]. As a function of m, it is determined as:

$$Q(m) = n(n+2) \sum_{k=1}^{h} \frac{r_k^2}{n-k} \tag{2}$$

where n is the number of data points.

The null hypothesis for LjungBox test states that the data is independently distributed (any observed correlations in the data is a result of the randomness of the sampling process). The alternate hypothesis is that the data is not independently distributed; it exhibits autocorrelation. The p-value is used to

[4] The autocorrelation was checked up to seven lags using the PerformanceAnalytics library in R[32].

	r_{daily}	r_{hourly}	$r_{7am-8pm}$
Reality Mining	0.55	0.89	0.65
Paki-Smartphone	0.15	0.60	0.44

Table 1. Proportion of ego-alter pairs demonstrating significant autocorrelation at different time periods.

decide if data points are not independently distributed. Typically, when the p-value is less than 0.05, the null hypothesis for LjungBox test is rejected. For each ego-alter pair, an hourly and a daily autocorrelation measure was calculated using the LjungBox test where a $p-value < 0.05$ means there is autocorrelation. Table 1. lists the proportion of ego-alter pairs that displayed autocorrelation in each of the two datasets. As the hourly autocorrelation measure may have been biased by lack of activity at night hours, a third autocorrelation for communication events between 7am and 8pm was also calculated. For each ego, we removed the alters who were communicated less than the mean of the communication frequency for that ego. This was plausible since on average, the mean lied in the upper quantile (with $\sigma > \mu$) of the calling distribution. The number of top contacts ranged between 5 and 20. This way we also removed the sparse data having insufficient number of communication events required to determine the autocorrelation.

As compared to our dataset, a higher percentage of ego-alter pairs in the reality mining dataset exhibit a daily as well as hourly periodic calling behaviour. The Reality Mining dataset was collected almost a decade ago when other means of smartphone communication such as Whatsapp, Viber, Facetime, etc. did not exist. In the smartphone dataset, we find that a small percentage of ego-alter pairs exhibit daily temporal autocorrelation. This might be an artifact arising from the shift in communication from mobile phone calls/text messages to smartphone instant messengers. Another tenable explanation could be the bias in the datasets. Contrary to the Reality Mining dataset that contains data from students or faculty of MIT media lab with daily activities structured around the academic calendar, the smartphone dataset contains data from general population of a developing country. Further, this is also an indication that communication patterns in different parts of the world may vary which is also acknowledged in [7] and thus it justifies the need to study the understudied populations that have a significant mobile phones user base. Notwithstanding a low proportion of time series exhibit autocorrelation in the daily interaction of Paki-Smartphone data, there is indeed an indication of periodic calling at finer levels.

CONCLUSIONS AND FUTURE WORK

In this paper we analyzed call data of Pakistani users which is an understudied population. Encouraged by our initial findings, we launched an android app and collected data from general population for a large scale analysis. On an aggregate level our data statistics were surprisingly comparable with the ones from Reality Mining [12] and Bentley and Chen [7]. We analyzed daily and weekly temporal patterns, showed that distribution of calls in an ego profile follows the rank size rule, detected periodicity at ego-alter pair level using autocorrelation and compared the results with the Reality Mining dataset. Further, we empirically observed that the present call logs scheme i.e., ordering of previous calls in reverse chronological order is a reasonably efficient way of dialing future calls. We also deliberated on the rationale behind high percentage of missed calls in our dataset.

Our results imply:

1). Call logs are an effective way of dialing future calls; 2). A comparison of datasets from high income vs. low income countries may further deepen our understanding about usage of mobile phones in different circumstances; 3). A reasonable number of ego-alter interactions do have temporal patterns at different time scales.

In future we would like to see whether our results can be replicated if we take a larger representative sample that can be generalized to all mobile phone users, especially those in the lower income countries. Another possibility is an attempt to identify daily and hourly patterns for ego-alter pairs, through other periodicity detection methods, besides autocorrelation; if this can be done then one could redesign the calling interface for mobile phones and improve the user experience significantly. The importance of intuitively sorting communication events (on displays) entails constant improvement in the user interface of interactive products, services, or systems. Such an interface, theoretically, would know the most likely people one is going to call at a given time and day. As a next step, we would like to study how users respond to an improved call log interface.

ACKNOWLEDGMENTS

Mehwish Nasim was funded by DEUTSCHE FORSCHUNGSGEMEINSCHAFT(DFG) under grant Br 2158/6-1.

APPENDIX A

Data statistics from the survey questionnaire:

Demographic data: Out of 28 participants, 13 were males and 15 were females. **Survey Responses:** 1).Would you agree with the statement “I have contact with different people on weekends as opposed to weekdays”: Strongly Agree(7%), Agree(42%), Neutral(25%), Disagree(25%). 2).Would you agree with the statement “I have experience in the usage of missed call as an indication for the other person to call back”: Strongly Agree(15%), Agree(46%), Neutral(21%), Disagree(14%), Strongly Disagree(3%). 3). How often do you use call log to dial a number from your cell phone: Always(25%), Usually (46%), About 50% of the time, (21%), Rarely (7%).

Data statistics from the preliminary call log dataset:

Demographic data: Out of 13 participants, 7 were males and 6 were females. All the participants were from Islamabad Capital Territory, Pakistan. **Call Statistics:** 4682 total calls; 19% missed calls; 54% outgoing calls; 27% incoming calls and remaining 1% were rejected calls. One participant had only outgoing calls in the dataset. Data had timestamps between 01 February 2015 and 31 March 2015. **Education Level:**M.Sc (Electronics): 2; Below High School: 4;

B.E (Chemical Engineering):2; M.S (System Engineering):2; High School:2; Bachelor of Arts: 1.

REFERENCES

1. Aileen Agüero and Harsha De Silva. 2009. Bottom of the pyramid expenditure patterns on mobile phone services in selected emerging Asian countries. In *4th Communications Policy Research, South Conference, Negombo, Sri Lanka*.
2. Talayeh Aledavood, Sune Lehmann, and Jari Saramäki. 2015a. On the Digital Daily Cycles of Individuals. *arXiv preprint arXiv:1507.08199* (2015).
3. T. Aledavood, E López, S. G. B. Roberts, F. Reed-Tsochas, E. Moro, R. I. M. Dunbar, and J. Saramäki. 2015b. Channel-Specific Daily Patterns in Mobile Phone Communication. *ArXiv e-prints* (July 2015).
4. T. Aledavood, E. López, S. G. B. Roberts, F. Reed-Tsochas, E. Moro, R. I. M. Dunbar, and J. Saramäki. 2015c. Daily rhythms in mobile telephone communication. *ArXiv e-prints* (Feb. 2015).
5. Albert-Laszlo Barabasi. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435, 7039 (2005), 207–211.
6. Osama O Barzaiq and Seng W Loke. 2011. Adapting the mobile phone for task efficiency: the case of predicting outgoing calls using frequency and regularity of historical calls. *Personal and Ubiquitous Computing* 15, 8 (2011), 857–870.
7. Frank Bentley and Ying-Yu Chen. 2015. The Composition and Use of Modern Mobile Phonebooks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2749–2758.
8. Ofer Bergman, Andreas Komninos, Dimitrios Liarokapis, and James Clarke. 2012. You never call: Demoting unused contacts on mobile phones using DMTR. *Personal and Ubiquitous Computing* 16, 6 (2012), 757–766.
9. Ulrik Brandes, Garry Robins, Ann McCranie, and Stanley Wasserman. 2013. What is network science? *Network Science* 1, 01 (2013), 1–15.
10. Julián Candia, Marta C González, Pu Wang, Timothy Schoenharl, Greg Madey, and Albert-László Barabási. 2008. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical* 41, 22 (2008), 224015.
11. Alessio Cardillo, Giovanni Petri, Vincenzo Nicosia, Roberta Sinatra, Jesús Gómez-Gardeñes, and Vito Latora. 2014. Evolutionary dynamics of time-resolved social interactions. *Physical Review E* 90, 5 (2014), 052825.
12. Nathan Eagle and Alex Pentland. 2006. Reality mining: sensing complex social systems. *Personal and ubiquitous computing* 10, 4 (2006), 255–268.
13. Nathan Eagle, Alex Sandy Pentland, and David Lazer. 2009. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences* 106, 36 (2009), 15274–15278.
14. Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. 2007. Youtube traffic characterization: a view from the edge. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 15–28.
15. Mohamed Ramzi Haddad, Hajer Baazaoui Zghal, Djemel Ziou, and Henda Ben Ghézala. 2014. A predictive model for recurrent consumption behavior: An application on phone calls. *Knowl.-Based Syst.* 64 (2014), 32–43. DOI: http://dx.doi.org/10.1016/j.knosys.2014.03.018
16. Marina Hennig, Ulrik Brandes, Stephen P Borgatti, Jürgen Pfeffer, and Ines Mergel. 2012. *Studying social networks: A guide to empirical research*. Campus Verlag.
17. Goolge Inc. 2010. Adaptive contact list. (14 09 2010). https://www.google.com/patents/US7797293
18. Zhi-Qiang Jiang, Wen-Jie Xie, Ming-Xia Li, Boris Podobnik, Wei-Xing Zhou, and H Eugene Stanley. 2013. Calling patterns in human communication dynamics. *Proceedings of the National Academy of Sciences* 110, 5 (2013), 1600–1605.
19. Hang-Hyun Jo, Márton Karsai, János Kertész, and Kimmo Kaski. 2012. Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics* 14, 1 (2012), 013055.
20. Andreas Kaltenbrunner, Vicenç Gómez, Ayman Moghnieh, Rodrigo Meza, Josep Blat, and Vicente López. 2007. Homogeneous temporal activity patterns in a large online communication space. *arXiv preprint arXiv:0708.1579* (2007).
21. Hayang Kim, Hui Zang, and Xiaoli Ma. 2013. Analyzing and modeling temporal patterns of human contacts in cellular networks. In *Computer Communications and Networks (ICCCN), 2013 22nd International Conference on*. IEEE, 1–7.
22. Gautier Krings, Márton Karsai, Sebastian Bernhardsson, Vincent D Blondel, and Jari Saramäki. 2012. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science* 1, 4 (2012), 1–16.
23. Conrad Lee, Bobo Nick, Ulrik Brandes, and Pádraig Cunningham. 2013. Link prediction with social vector clocks. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 784–792.

24. Amanda Lenhart. 2010. *Cell phones and American adults*. Technical Report. Pew Research Center. http://www.pewinternet.org/2010/09/02/cell-phones-and-american-adults/

25. Greta M Ljung and George EP Box. 1978. On a measure of lack of fit in time series models. *Biometrika* 65, 2 (1978), 297–303.

26. mmaglobal. 2015. Mobile: The Closest You Can Get to Your Consumers. (2015). http://www.mmaglobal.com/case-study-hub/case_studies/view/36713/ Accessed: 2016-02-05.

27. Mehwish Nasim and Ulrik Brandes. 2014. Predicting Network Structure Using Unlabeled Interaction Information. In *MMB & DFT 2014 Proceedings of the International Workshop SOCNET 2014*. University of Bamberg Press, 57–64.

28. Mehwish Nasim, Muhammad Usman Ilyas, Aimal Rextin, and Nazish Nasim. 2013. On commenting behavior of Facebook users. In *24th ACM Conference on Hypertext and Social Media*. ACM, 179–183.

29. Mehwish Nasim, Aimal Rextin, Numair Khan, and Muhammad Muddassir Malik. 2015. On Temporal Regularity in Social Interactions: Predicting Mobile Phone Calls. *arXiv preprint arXiv:1512.08061* (2015).

30. Antti Oulasvirta, Mika Raento, and Sauli Tiitta. 2005. ContextContacts: re-designing SmartPhone's contact book to support mobile awareness and collaboration. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*. ACM, 167–174.

31. Srinivasan Parthasarathy, Sameep Mehta, and Soundararajan Srinivasan. 2006. Robust periodicity detection algorithms. In *Proceedings of the 15th ACM international conference on Information and knowledge management*. ACM, 874–875.

32. Brian G. Peterson and Peter Carl. 2014. PerformanceAnalytics: Econometric tools for performance and risk analysis. (2014). http://CRAN.R-project.org/package=PerformanceAnalytics R package version 1.4.3541.

33. Santi Phithakkitnukoon, Ram Dantu, Rob Claxton, and Nathan Eagle. 2011. Behavior-based adaptive call predictor. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 6, 3 (2011), 21.

34. PTA. 2016. Telecom Indicators. (2016). http://www.pta.gov.pk/index.php?option=com_content&view=article&id=269&Itemid=599/ Accessed: 2016-02-08.

35. Kenneth T Rosen and Mitchel Resnick. 1980. The size distribution of cities: an examination of the Pareto law and primacy. *Journal of Urban Economics* 8, 2 (1980), 165–186.

36. Andres Sevtsuk and Carlo Ratti. 2010. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology* 17, 1 (2010), 41–60.

37. Lionel Tabourier, Anne-Sophie Libert, and Renaud Lambiotte. 2016. Predicting links in ego-networks using temporal information. *EPJ Data Science* 5, 1 (2016), 1–16.

38. M ten Thij, S Bhulai, and P Kampstra. 2014. Circadian patterns in twitter. *DATA ANALYTICS* (2014), 12–17.

39. Theophanis Tsandilas and others. 2005. An empirical assessment of adaptation techniques. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2009–2012.

40. Ioannis Tsaousis. 2010. Circadian preferences and personality traits: A meta-analysis. *European Journal of Personality* 24, 4 (2010), 356–373.

41. Rütger A Wever. 2013. *The circadian system of man: results of experiments under temporal isolation*. Springer Science & Business Media.

42. Wikipedia. 2016. List of countries by number of mobile phones in use. (2016). https://en.wikipedia.org/wiki/List_of_countries_by_number_of_mobile_phones_in_use/ Accessed: 2016-02-08.

43. Ye Wu, Changsong Zhou, Jinghua Xiao, Jürgen Kurths, and Hans Joachim Schellnhuber. 2010. Evidence for a bimodal distribution in human communication. *Proceedings of the national academy of sciences* 107, 44 (2010), 18803–18808.

44. Taha Yasseri, Robert Sumi, and János Kertész. 2012. Circadian patterns of wikipedia editorial activity: A demographic analysis. *PloS one* 7, 1 (2012), e30091.

45. Eviatar Zerubavel. 1985. *Hidden Rhythms: schedules and calendars in social life*. Univ of California Press.

46. Zui Zhang, Hua Lin, Kun Liu, Dianshuang Wu, Guangquan Zhang, and Jie Lu. 2013. A hybrid fuzzy-based personalized recommender system for telecom products/services. *Information Sciences* 235 (2013), 117–129.