Pervasive and Mobile Computing 41 (2017) 166–178

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Data analysis and call prediction on dyadic data from an understudied population

Mehwish Nasim [a,b,*], Aimal Rextin [c], Shamaila Hayat [c,d], Numair Khan [e], Muhammad Muddassir Malik [f]

[a] *Department of Computer & Information Science, University of Konstanz, Germany*
[b] *School of Mathematical Sciences, The University of Adelaide and ARC Centre of Excellence for Mathematical and Statistical Frontiers, Australia*
[c] *Department of Computer Science, COMSATS Institute Of Information Technology, Islamabad, Pakistan*
[d] *Department of Computer Science, The University of Poonch, Rawlakot, Azad Jammu and Kashmir*
[e] *Department of Computer Science, Brown University, USA*
[f] *National University of Sciences and Technology, Pakistan*

A R T I C L E I N F O

Article history:
Received 28 March 2017
Received in revised form 7 July 2017
Accepted 12 August 2017
Available online 24 August 2017

Keywords:
Smartphone
Call-logs
Call prediction
Temporal regularity

A B S T R A C T

In this paper we predict outgoing mobile phone calls using machine learning and time clusters based approaches. We analyze to which extent the calling activity of mobile phone users is predictable. The premise is that mobile phone users exhibit temporal regularity in their interactions with majority of their contacts. In the sociological context, most social interactions have fairly reliable temporal regularity. If we quantify the extension of this behavior to interactions on mobile phones we expect that pairwise interaction is not merely a result of randomness, rather it exhibits a temporal pattern. To this end, we not only tested our approach on an original mobile phone usage dataset from a developing country, Pakistan, but we also analyzed the famous Reality Mining Dataset and the Nokia Dataset (from a European country), where we found an equitable basis for comparison with our data. Our original data consists of 783 users and more than 12,000 active dyads. Our results show that temporal information about pairwise user interactions can predict future calls with reasonable accuracy.

1. Introduction

Social networks are made up of a set of social entities (people, actors, organizations etc.) and social relations (friendship, kinship, etc.), between those entities. Social relations consists of persistent relations such as friendship and instantaneous relations (interactions) such as talk to, joint participation in an event, extend help to, etc. Seemingly autonomous individuals and organizations in a social network are, in fact, embedded in social relations and interactions. Massive amount of relational event data is generated by social interactions. Such data, as proxy of human relationships is helpful in understanding and predicting behavior of individuals such as influence, activity bursts, buying habits etc.

Mobile phones are the most common medium for social interactions. In America alone, there are almost 1.3 billion mobile communication events daily [1]. Because of mobile phone's portable nature, a communication event can take place in a variety of situations and one can assume very little about the context of a call.

* Corresponding author at: School of Mathematical Sciences, The University of Adelaide and ARC Centre of Excellence for Mathematical and Statistical Frontiers, Australia

E-mail addresses: mehwish.nasim@adelaide.edu.au (M. Nasim), aimal.rextin@comsats.edu.pk (A. Rextin), shamailahyt@yahoo.com (S. Hayat), nkhan6@cs.brown.edu (N. Khan), muddassir.malik@seecs.edu.pk (M.M. Malik).

http://dx.doi.org/10.1016/j.pmcj.2017.08.002
1574-1192/

This work is motivated by two factors: (1) User interactions on social media such as on smartphones carry important information about the underlying social dynamics. Discovering this knowledge is challenging because of the frequency and versatility of use of smartphones. (2) Further, the same two factors also necessitate an extremely efficient call-making interface design.

Contributions:

1. Most of the call-logs data analyzed in the literature is from developed countries [2]. We collected an original mobile phone usage dataset of 783 users with 229,450 communication events from an understudied population: Pakistani mobile phone users. This is one of the potential strengths of our work.
2. We explore temporal homogeneity/non-homogeneity in mobile phone calls, in order to predict future communication events between pairs of individuals. We perform a study of possible features in time series analysis that are useful in call prediction. Using actual call logs we show that majority of users are not optimally served by existing calling applications such as call logs. Further, we also test the hypothesis that, majority of caller–callee interactions display temporal regularity through a statistical measure called autocorrelation. We then propose a machine learning call prediction method based on temporal regularities between ego-alter pairs and perform experiments on both the collected data as well as on the famous Reality Mining Dataset [3] to demonstrate applicability of our methods for predicting future calls.
3. Further, we show that most ego-alter pairs call around the same time and use this observation to propose a call prediction algorithm and compare it with algorithm proposed by Stefanis [4] on their dataset [5].

2. 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 [6–8]. Activity on twitter in various languages shows that circadian patterns exist for tweets all around the world [9]. 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) [10]. Temporal interactions have also been used to predict links in social networks [11–13].

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 [14]. Data of calling patterns has been used to infer friendships relations and uncover individual and collective human dynamics [3,14–18]. 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 [19]. Inspired by effective studies on calling patterns, researchers have devised several call prediction models. In [20], authors predicted the outgoing and incoming calls on Reality Mining dataset [3] 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. [21] modeled the historic call patterns of users and achieved a 35% accuracy for call prediction on synthetic data. Haddad et al. [22] 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 [23–25]. Barabasi [23] attributes this bursty non-Poisson character of human behavior to a priority-based queuing process. This view is supported by Jo et al. [24] 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. [26] 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. [27] 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 [28]. While studying the effects of two different UI adaptation techniques on user performance, Tsandilas and Schraefel [29] conclude that adaptation is always more effective, even when the accuracy of prediction is low. Bentley and Chen [30] 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. In addition, they found that a median of $Q = 60\%$ (six out of a total of $N = 10$ contacts) displayed in a "recent calls" list are amongst the most frequently contacted. While the authors use this latter statistic to argue against the efficacy of a "recent calls" list, it would be interesting to explore whether the value of Q increases for larger values of N, especially since the authors' results indicate an upward trend in Q as N increases from one to 10. Proponents of a "recent calls" list may argue that, in practice, these lists hold more than 10 entries. Based on their results, Bentley and Chen suggest a redesign of the content and representation of contact lists. A redesign of contacts book was proposed also in [31]. 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.

In social network settings, 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) [10]. Temporal interactions have also been used to predict links in social networks [11–13,32]. 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 [6–8]. Activity on twitter in various languages shows that circadian patterns exists for tweets all around the world [9]. While studying social network turnover, Aledavood et al. [33,34] 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. [33–35] are 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.

Systems with time-stamped dyadic[1] interactions can be modeled as temporal networks. Time stamped networks of human communication along with proximity networks, are the largest class of systems modeled as temporal networks. When the dependence between interaction values in the past is preserved in the future, then future interactions are reasonably easy to predict.

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 [36]. 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. In numerous studies [24,26] inhomogeneity has also been observed in human activities. [34] showed that the individual differences in the distribution of calling remain persistent. They also suggested that frequently called contacts are the ones most likely to be contacted during low entropy periods. Furthermore, causal events are also a key characteristic of human communication behavior. [37] showed that communication behavior varies between pairs of users. All these characteristics make the prediction of future interactions more challenging.

3. Exploratory data analysis

This work is a continuation of our earlier work [37] where we provided some statistics about our data. We collected data of 783 users, with 229,450 communication events for analysis, called the Smartphone dataset. Section 3.1 summarizes our earlier work, and Section 3.2 extends the data analysis by testing for autocorrelation, entropy and time based clusters.

Our dataset consists of 24% incoming and answered calls, 19% incoming and missed calls, and 54% of calls were outgoing calls. Bentley and Chen [30] also reported similar statistics for their dataset of 200 users. These statistics are also comparable with the statistics of Reality Mining dataset, with slight variations in the percentage of missed and outgoing calls. A relatively high percentage of missed calls is a noticeable artifact of our dataset. In a low income country, missed calls are used to indicate some signal, which is an easy way to save money.

Moreover, we observe that, only a small fraction of contacts are called frequently. Bentley [30] observed that most calls are to 5–10 of contacts. Miritello et al. [38] we observed that individuals exhibit a finite communication capacity, which limits the number of ties they can maintain. Similarly, [39] observed that people normally do 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 the value of a varied between 0 and 7 and of b varied between 0 and 2.5. We observed that our equation fits the data very well and we got a mean adjusted R^2 of 0.89 with a standard deviation of 0.16 in Smartphone dataset. We also checked the results on the dataset used by [4], the Nokia dataset [5]. We got mean adjusted R^2 of 0.94 and the standard deviation was 0.02. The typical shape of call distribution curve for an ego[2] can be seen in Fig. 1. We

1 dyad: pair of users.

2 ego is the focal actor and alters are his/her contacts.

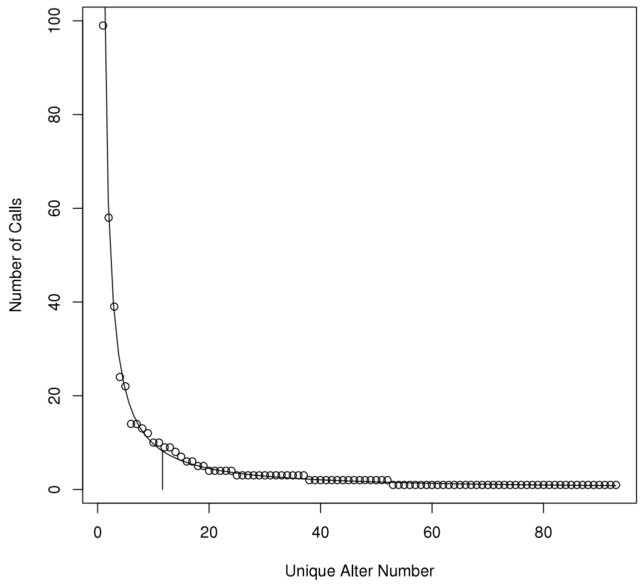


Fig. 1. Calls distribution for a particular ego.

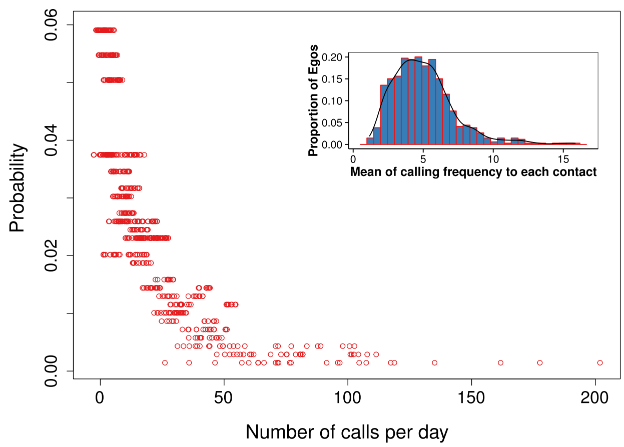


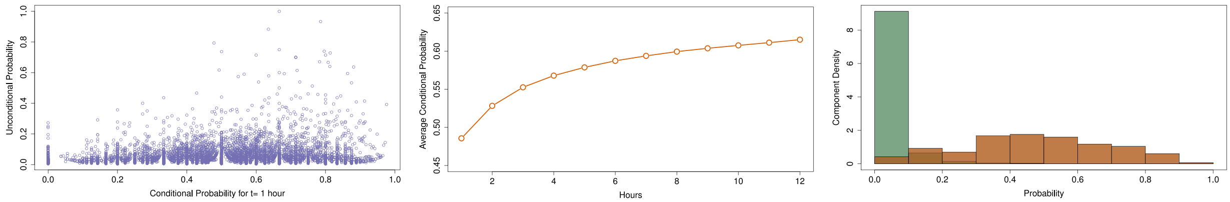
Fig. 2. Distribution of calls.

find the point of bend, by equating the differential of Eq. (1), to 45^o. We think, the most important contacts of an ego lie above this point. 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* [40]. 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 Eq. (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 Fig. 2. On average, each user made or received ≈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.

3.1. Calling behavior of ego-alter pair

In this section, we discuss the difference in communication behavior of ego-alter pairs on weekends and weekdays, and variation in conditional probability of a communication event between an ego-alter pair. We also test temporal regularity for caller–callee interactions by applying different statistical tests.

Regarding difference in probability of calling an alter, we analyzed 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 the difference in calling behavior is probably due to the nature of social relation of ego-alter pairs.



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

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

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

Fig. 3. Probability of communication. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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

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

Moreover, an informal user analysis (using a questionnaire) showed that 71% of respondents either always or usually use call log to initiate a call [37]. If this trend is true in general, then 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 $t \in \{1, 2, \ldots, 12\}$. Fig. 3a clearly showed that a communication event is much more likely if there was a communication event within the last 60 min. However, this conditional probability does not increase significantly and levels off very quickly as we increase t, as shown in Fig. 3b. Moreover, Fig. 3c shows 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 deviation ($L(\mu, \sigma^2)$) of this distribution using the Maximum Likelihood Estimate. We estimated $\mu = 0.48$ and $\sigma^2 = 0.21$ as compared to the unconditional probability where $\mu = 0.041$ and $\sigma^2 = 0.059$. To the best of our knowledge, we interpret these results as the first empirical evidence that existing calling interfaces such as call logs are fairly useful when users want to make a call. Nonetheless, for a certain set of ego-alter pairs there is still room for improvement.

We also tested for the correlation between the conditional calling probability and number of communication events. Our results suggest significant positive correlation between the two variables.

3.2. Temporal regularities in caller–callee interactions

We applied different statistical tests to analyze the temporal regularities in caller–callee interactions. The results of these tests are summarized below.

Autocorrelation: We test the hypothesis that the majority of caller–callee interactions display temporal regularity. Formally, we state the hypotheses as follows:

The alternative hypothesis (H_A) states that *the proportion of ego-alter pairs displaying periodic communication patterns is* $\geq$*50%*. This is based on the assumption that the majority, i.e., more than half, of all communication events are periodic. The null hypothesis (H_0), consequently, is that *the proportion of ego-alter pairs that display periodic communication patterns (autocorrelation) is* <50% .

For each ego-alter pair in both datasets, an hourly and a daily autocorrelation measure was calculated using the Ljung Box 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.

Based on the results for the Reality Mining dataset, we reject the null hypothesis for both the daily (p<2.2×10^{-16}), and the hourly autocorrelation measures ($p < 2.2 \times 10^{-16}$). This implies that more than 50% of ego-alter pairs in the Reality Mining dataset demonstrate periodic calling behavior at the daily and daytime-hours granularity level. For the Smartphone dataset, we fail to reject the null hypothesis for the daily autocorrelation measure ($p = 1$). However, we reject the null hypothesis for the hourly ($p < 2.2 \times 10^{-16}$) and the daytime-hours ($p < 2.2 \times 10^{-16}$) autocorrelation measures.

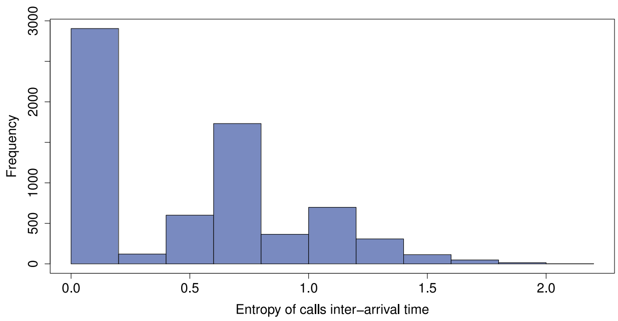
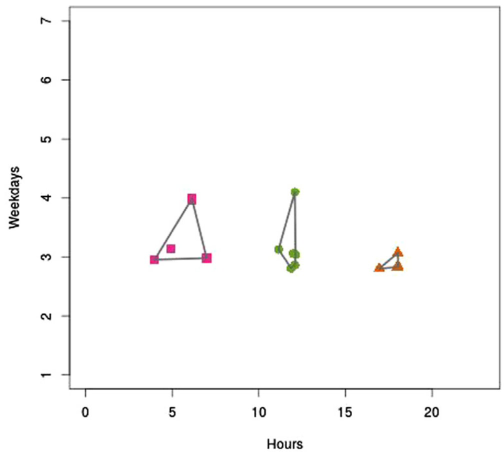
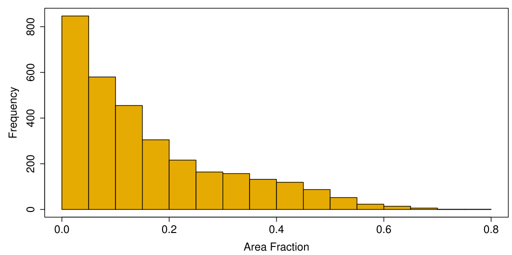


Fig. 4. Entropy of calls inter-arrival time.



(a) 2D plot showing a weekly and hourly calling activity in a particular ego-alter pair.



(b) Histogram showing area fractions and their frequency.

Fig. 5. Histogram and 2D plot showing the weekly and hourly temporal patterns between caller–callee interactions.

Entropy: We also examined the entropy of inter-arrival time for all call pairs. A higher entropy would mean there is less regularity in the calling behavior while lower entropy points periodicity or burstiness. An ego-alter pair who talk around the same time everyday would have an entropy value of zero.

We divided the time into bins, each corresponding to one hour. We then created a vector of counts for each bin and calculated the entropy. Fig. 4 shows that about 3000 ego-alter pairs have an entropy value of zero. About 45% of all ego-alter pairs have an entropy value of less than 0.5. This is consistent with the results that we obtained for the autocorrelation metric, where 44% ego-alter pairs exhibited autocorrelation in their hourly communication.

Time Clusters: We first establish the rationale behind the selection of dimensions (features) for predicting future calls. We assume that people mostly call a particular contact at a specific time of the day, e.g. in the afternoon or in the evening etc. To prove this assumption true, we conducted a pretest where for each *ego-alter* pair with at least 15 communication events, we extracted the *hour of the day* and *day of the week* for each communication event and accordingly plotted it on a 2D space.

We then used the *DBSCAN clustering* algorithm to find clusters in these communications events and then computed the convex hull around each of these clusters as shown in Fig. 5a. The intuition is that small polygon like clusters indicate that the dyadic (ego-alter) communication has a temporal component, whereas, communication events dispersed over the plot indicate non-homogeneity in communication. Fig. 5b, represents a histogram, showing the area fractions of different clusters and their frequency. It is more clear from the histogram that small fractional areas are more than large areas, which is a clear indication of temporal regularity. In fact, it turned out that of 3162 dyads, in 3064 cases this area is less than 0.5 whereas in 98 pairs, this area is between 0.5 and 0.7. This indicates that a calling pattern exists in most of the ego-alter communication.

We analyzed the relationship between number of communication events and calling regularity in Fig. 6. For time clusters, we computed the *Kendall rank correlation* on the number of communication events between an ego-alter pair and the *area fraction*. Kendall rank correlation is a non-parametric test which measures the strength of dependence between two

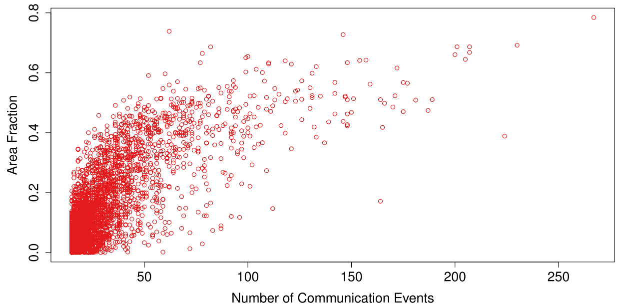


Fig. 6. Correlation between number of communication events and area fraction of time based clusters.

variables. We reject the null hypothesis that variables are uncorrelated at 0.05 significance level, since we obtained a p value of 2×10^{-16}. However, for autocorrelation and entropy we could not find any correlation between the number of communication events and value of the metrics.

4. Call prediction

In the previous section, we conducted an exploratory data analysis in order to identify features for call prediction. In this section, we propose two call prediction algorithms: one is based on machine learning techniques and the other is based on time clusters in communication events.

4.1. Machine learning algorithm

We classified our data using the Support Vector Machines (SVM) classifier, using the implementation available in R [41]. The explanatory features used for every call are: time of the day (correct to the nearest minute), weekday (Sunday–Saturday), Night-call (true or false) and direction of the call (incoming, outgoing or missed). We divided the data (each ego's call log) into training and test sets. We use 80% of the data for training the model and the remaining 20% data for predicting future calls[3] . We used a linear combination of the calling features for prediction. For every call in the test set, the classifier outputs probabilities against each class.

We observed that a few contacts are called more often. Hence, it is reasonable to remove those contacts which are sporadically contacted. For our experiments, we selected the mean of calling frequency as the cut off threshold. Among various types of contacts in the contact list, only few are called most often and rest are called rarely [26,39]. In the final analysis 10,383 caller–callee pairs are analyzed in the Smartphone dataset and 1851 pairs are analyzed in the Reality Mining dataset.

Evaluation metric

We have used the following two evaluation criteria for performance evaluation:

1. We have compared the performance of our approach with top-k most frequently called numbers and with last-k calls. In a hypothetical situation, whenever a user presses the call button or opens the calling interface, at time t_n, a list of contacts is displayed. Our classifier outputs probabilities for each class (contact) a user is likely to call at time t_n. These probabilities are computed and sorted and a list of contact numbers with the highest probabilities is displayed. We call this list, 'top-k recommendations'. We calculate the probability that u_i is going to call $x_j \in X_i$ (or vice versa), given that θ_j amount of time has elapsed since the last communication. We denote this probability by $P(x_j|\theta_j)$. We observed that when θ_j is small or when the last communication event was a missed call from x_j (or to x_j), the probability to communicate with x_j is high. For a give θ_j, we pick the last-k' calls and include them in the results that we obtained from our classifier. We then generate a final list of most likely numbers to be dialed at any given time (within the next hour) based on the results of the classifier and last-k' calls.
2. In the second evaluation method, we measure the proportion of calls that are predicted within a certain error threshold (ϵ). For a given time, a 'single phone number' is predicted which the user is likely to call. We then measure how well the number is predicted with regards to different time-deviation thresholds.

[3] The Smartphone dataset contains data from April till September 2015. On average, the training set contained data approximately from April–August 2015 (about 16 weeks), used to predict calls made between August–September 2015 (about 5 weeks). On average, training the model took 48.35 ms for each ego, on a Lenovo X1 Carbon Notebook with Intel Core i-7 CPU(2 GHz) and 8 GB of RAM.

Table 2
Proportion of correctly predicted calls in Reality Mining dataset for various list lengths for prediction deviation, $\epsilon = 1$ h. Here last-k' are the number of last calls used in the final list.

k	last-$k' = 2$	last-$k' = 3$
1	0.44	0.44
2	0.77	0.77
3	0.78	0.77
4	0.79	0.78
5	0.80	0.78
6	0.82	0.79
7	0.84	0.80
8	0.84	0.81
9	0.85	0.82
10	0.86	0.83

Table 3
Proportion of correctly predicted calls in Smartphone dataset for various list lengths for prediction deviation, $\epsilon = 1$ h. Here last-k' are the number of last calls used in the final list.

k	last-$k' = 2$	last-$k' = 3$
1	0.31	0.31
2	0.73	0.73
3	0.73	0.77
4	0.74	0.78
5	0.75	0.78
6	0.77	0.79
7	0.79	0.80
8	0.80	0.81
9	0.82	0.82
10	0.83	0.83

Table 4
Average accuracy (%) with different methods.

	Reality Mining			Smartphone		
	$k = 5$	$k = 10$	$k = 15$	$k = 5$	$k = 10$	$k = 15$
top-k called numbers	60.04	70.20	77.65	45.65	63.73	74.51
last-k numbers	63.78	69.59	73.58	63.94	69.76	72.83
top-k recommendations	80.70	86.66	88.46	74.96	83.72	84.08

Predictions are made for the users who have at least 50 communication events in the dataset. Hence we analyzed 89 users in the Reality Mining Dataset and 604 users in the Smartphone dataset. Further, we wanted to improve accuracy using fewer dimensions. For the last calls related features, we have used data pertinent to only the last two calls since there is a trade-off between adding dimensions to the feature set and efficiency.

Top-k recommendations

From the users' perspective, top-k recommendations should be more accurate as compared to last-k calls. We generate a list of most likely numbers to be dialed at any given time: the 'top-k recommendations'. We compare the accuracy of top-k recommendations with the accuracy obtained by last-k calls. We show the performance of our approach for individual users for varying list lengths(1 to 10). Table 4 reports the average performance of our approach along with performance achieved by baseline methods.

In Figs. 7 and 8, x-axis represents the users (egos) in the dataset. For every user in the datasets, we report the accuracy achieved by top-k recommendations vs. last-k calls and top-k called numbers (most frequently called contacts). The accuracy is reported for each user in *Reality Mining dataset:* points in blue; and *Smartphone dataset:* points in red. A higher concentration of points below the identity line indicates that top-k recommendations has better performance against the respective method. Table 4 reports the average performance of our approach along with performance achieved by baseline methods.

Tables 2 and 3 report the proportion of correctly predicted calls for various list lengths.

Prediction deviation

From the service providers' perspective, accurate prediction of calls would enable them to predict users' behavior and predict periods of high usage which in turn would lead to better load balancing, hence, better service quality. The results

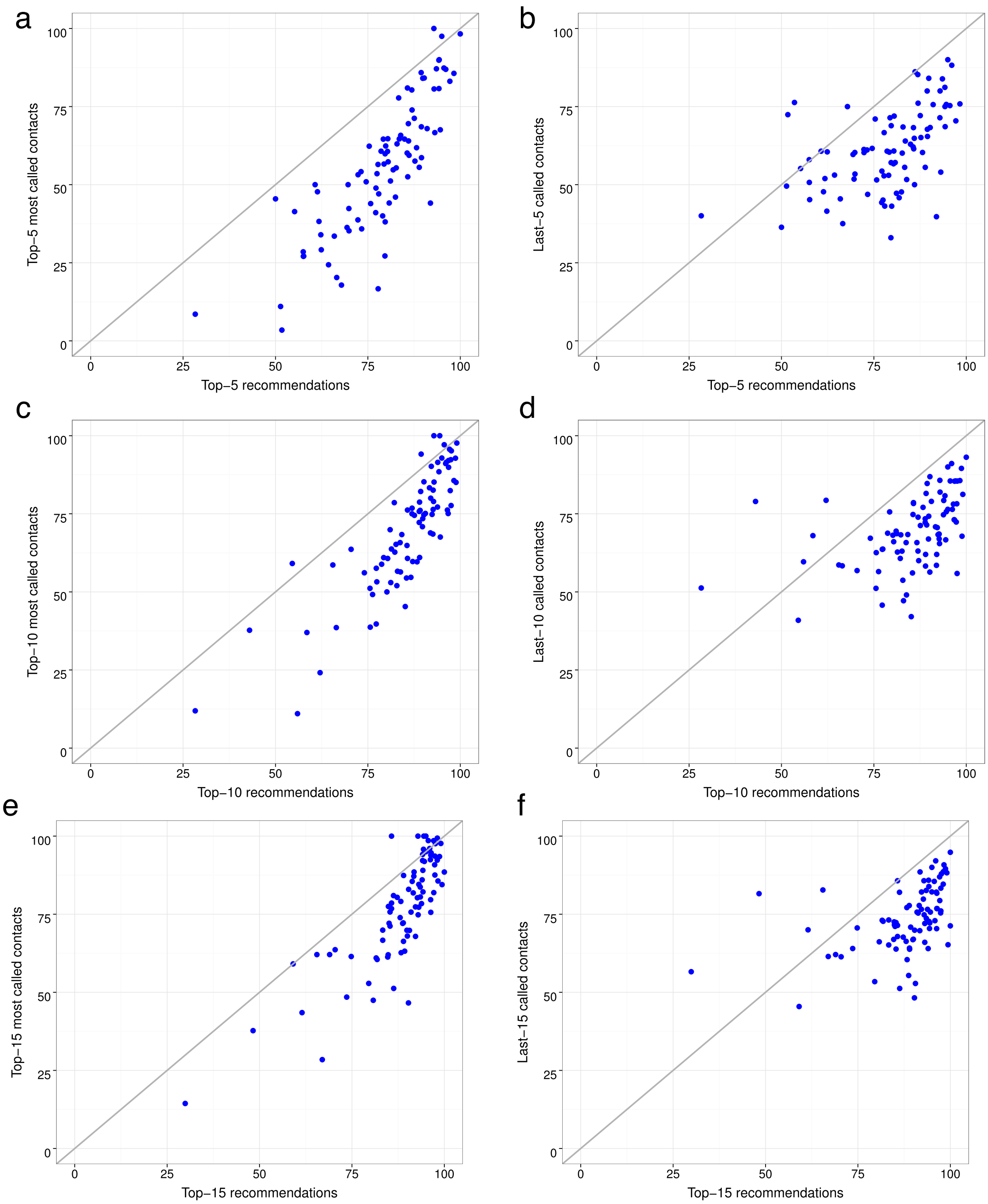


Fig. 7. These plots show accuracy of: top-k recommendations against top-k called numbers and last-k numbers for each user. Points below the identity line indicate that top-k recommendations has better performance against the respective baseline method. Performance is reported for: (a), (b) Reality Mining $- k = 5$. (c), (d) Reality Mining $- k = 10$. (e), (f) Reality Mining $- k = 15$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

show that a reasonable proportion of the phone calls are predictable using the proposed method. For the Reality Mining dataset 44% of the outgoing calls were predicted below one hour error threshold. For the Smartphone dataset 31% of the outgoing calls were predicted below one hour error threshold. Tables 2 and 3 report the proportion of correctly predicted calls for various list lengths for prediction deviation, $\epsilon = 1$ h.

4.2. Time clustering based call prediction algorithm

We earlier observed in Fig. 5a that calls between an ego-alter pair are concentrated in particular time regions. This indicates that ego-alter interactions follow certain temporal patterns. It also suggests that chances of interaction between

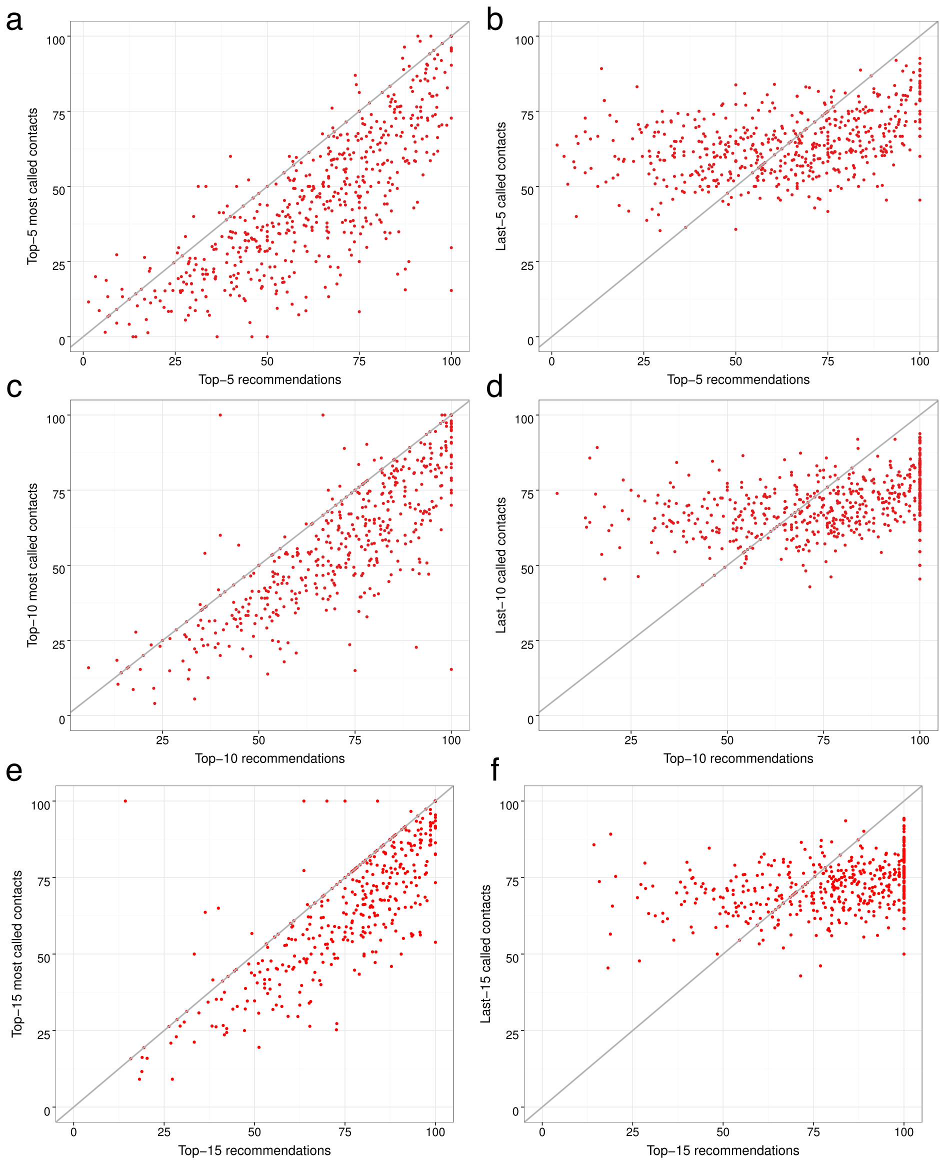


Fig. 8. These plots show accuracy of: top-k recommendations against top-k called numbers and last-k numbers for each user. Points below the identity line indicate that top-k recommendations has better performance against the respective baseline method. Performance is reported for: (a), (b) Smartphone — $k = 5$. (c), (d) Smartphone — $k = 10$. (e), (f) Smartphone — $k = 15$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

particular ego-alter pairs reduces as one moves away from these regions. This intuition is the basis of our algorithm that we discuss below:

Step 1: Identify clusters of each ego-alter pair and identify the convex hull for each cluster. We identified clusters by using the DBSCAN algorithm with default parameters. Since calls of each ego-alter pairs are generally divided into two or more regions, if a dyad has fewer calls then few clusters would be formed and there would be more incorrect predictions. Hence we needed to decide on a threshold value to filter out dyadic sets with low number of calls.

Step 2: The original polygon reflects the time and day an ego-alter pair is more likely to communicate with each other. Since human perception of time does not have strict boundaries, so we made two extended polygons to cater for the possibility of the ego-alter pair communicating at slightly different times. The additional polygons P_1 and P_2 are created around the original polygon and are of the same shape. The polygon P_1 has a size that is 1.5 times the size of the original polygon and the P_2 is double the size of the original polygon. A test call that falls in the original polygon is assigned a probability of 100%, while for P_1 and P_2 we assigned probabilities of 80% and 50% respectively.

Table 5
Correct prediction rate based on time clusters.

Comm. Events	Smartphone dataset			Nokia dataset		
	Correct predictions	Incorrect prediction	Avg.	Correct predictions	Incorrect predictions	Avg.
>0	20 563	24 413	46%	4424	1118	80%
>4	20 508	13 156	45%	4324	480	78%
>9	19 426	6 392	43%	3986	179	72%
>14	18 108	2 808	40%	3610	85	65%

We performed the evaluation of this algorithm by dividing the dataset into a training set consisting of 80% of the user's calls and a test set consisting of 20% of the user's calls. We calculated the probability of calling each alter using the above algorithm by using the earliest call in test set. We then selected the top probability alters to generate a top-k recommendations where the k is 8. The prediction was termed as a success if the alter of the call is in the top-k recommendation list, otherwise we considered it a failure. This call was then added to the training set and the same process was repeated for all calls in the test set one by one. We also experimentally tested other sizes and probabilities and found that the accuracy reduced at most by 4%. The accuracy of prediction is reported in Table 5.

5. Discussion

Mobile phones represent one of the most commonly used communication medium. The portable nature of the medium means very little can be assumed about the situation in which the phone is used; a typical user makes calls in all kinds of contexts. These two factors, frequency and versatility of use, necessitate an extremely efficient call-making interface design.

As a first step we conducted a pretest on two mobile phone datasets for determining whether users have a regular calling pattern or not. We modeled the communication between mobile phone users as a time series data analysis problem.

We collected and analyzed call data of Pakistani users, which is an understudied population. Call-logs and SMS datasets from understudied populations are rare [42]. Researchers have shown that different ethnic populations have distinct mobile phone usage characteristics. In their seminal work, Blondel et al. [43] analyzed mobile traffic of 2 million users in Belgium. They found that the two main ethnic groups in the country, i.e., Walloons and Flemish, can be clearly inferred from the mobile call graph. On an aggregate level our data statistics were surprisingly comparable with the ones from Reality Mining [3], and Bentley and Chen [30]. We analyzed daily and weekly temporal patterns, showed that distribution of calls in an ego profile follows the rank size rule, detected periodicity at dyadic level using autocorrelation and compared the results with the Reality Mining dataset. Further, we empirically observed that call logs are an efficient way of dialing future calls. We also deliberated on the rationale behind high percentage of missed calls in our dataset.

In many time series, it is plausible to expect that the recent data points are likely to have an influence on the future data points. In order to identify whether the ego-alters communication data has a pattern, we used autocorrelation which is a type of correlation statistic specifically for correlating the recent data point to other data points in the series. The results on two different datasets quantitatively show that a reasonable number of ego-alter pairs exhibit autocorrelation.

The results show that more than 50% of ego-alter pairs in the reality mining dataset exhibit a daily as well as hourly periodic calling behavior. 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 fail to find 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. Notwithstanding that a low proportion of time series exhibit autocorrelation in the daily interaction of Smartphone data, there is indeed an indication of periodic calling at finer levels.

We then predicted the calling behavior of mobile phone users (given the time based features), using a machine learning approach and using few dimensions. We have identified the day of the week and time as two important features, that help in accurately predicting the next outgoing call. This is supported by the fact that human interaction behavior follows a circadian rhythm. We have also analyzed the situations where it is more probable that the user calls a number from one of the last called numbers.

Predictive analytics deals with understanding the data, extracting information from data and using it to predict trends and patterns. Most often the event of interest is in the future (e.g., predicting links, buying behavior, etc.), but predictive analytics can also be applied to any type of unknown event. In predictive analytics, a feature is any important piece of information about the data that might be useful for the prediction task. The purpose of a feature, other than being an attribute, would be much easier to understand in the context of a problem. Although, machine learning methods are a disadvantage when requisite data cleansing has not been done, compared to time series forecasting techniques, by careful feature selection one may obtain reasonable results.

Similar evaluation methods for call prediction have been used in previous studies. With a few exceptions, most previous studies used different datasets for analyzing calling behavior, therefore, a direct comparison is not equitable. [20] predicted the outgoing and incoming calls on Reality Mining dataset. Out of the 94 users, they selected only 30 users for experiments. The identities of those users are not disclosed in the paper, therefore, a direct comparison with their results is not possible. For completion, we have reported the performance of our method on 89 out of 94 users. The remaining 5 users had less than 50 communication events. For outgoing call prediction, they also generated a list of most likely numbers to be dialed at any given time. For the 30 random users in their experiments they achieved an accuracy of 41% if the predicted list is only allowed one entry. If the predicted list has five entries their model correctly predicted the dialed number 70% of the time. On the Reality Mining dataset we achieve an accuracy of 44% when the top-k list has one entry. Our results show more than 78% accuracy on the Reality Mining dataset when the predicted list is allowed 5 entries. Table 4 shows that our approach also performs better than the last-k calls on both the datasets.

Authors in [21] modeled the historic call patterns of users and achieved a 35% accuracy for call prediction on a synthetic dataset. Haddad et al. [22], report the prediction accuracy for certain time-deviation thresholds on a dataset consisting of more than seven thousand users. Their model predicted about 17% of the outgoing calls with an error below one hour.

In the previous models such as the ones proposed in [22] and [20], a baseline comparison was missing. The motivation behind our study was to come up with a method that could better predict the next call. Hence, from the user's point of view we found it imperative to check the performance of the last-k calls as well. It is a reasonable expectation that a call prediction approach should perform better than the current approach used for smartphone call logs i.e., displaying the recent calls in chronological order. Table 4 shows that our approach performs better than the last-k calls and most frequently called numbers list. Our call prediction approach outperformed the two baseline approaches i.e. predicting next call based on last-k calls and predicting next call using the most frequently called numbers' list. We found it very intriguing as it opens many exciting research questions. One of them is to see whether these results can be replicated if we take a large representative sample that can be generalized to all mobile phone users. In order to deeply understand the phone call behavior, it is important to analyze a large call logs dataset along with other relevant information such as demographic, geographical, and socio-economic data. Another future research possibility could be an attempt to redesign the calling interface for mobile phones which could improve the user experience significantly. Such an interface, theoretically, would know the most likely people one is going to call at a given time and day. In future we would like to study how users respond to an improved call log interface.

Acknowledgments

Mehwish Nasim acknowledges the financial support from DEUTSCHE FORSCHUNGSGEMEINSCHAFT (DFG) under grant Br 2158/6-1.

Numair Khan contributed in this work while he was working at National University of Sciences and Technology, Pakistan as a research assistant.

References

[1] Amanda Lenhart, Cell phones and american adults, Technical report, Pew Research Center, 2010. http://www.pewinternet.org/2010/09/02/cell-phones-and-american-adults/.
[2] Diala Naboulsi, Marco Fiore, Stephane Ribot, Razvan Stanica, Large-scale mobile traffic analysis: a survey, IEEE Commun. Surv. Tutor. 18 (1) (2016) 124–161.
[3] Nathan Eagle, Alex Pentland, Reality mining: sensing complex social systems, Pers. Ubiquitous Comput. 10 (4) (2006) 255–268.
[4] Vassileios Stefanis, Athanasios Plessas, Andreas Komninos, John Garofalakis, Frequency and recency context for the management and retrieval of personal information on mobile devices, Pervasive Mob. Comput. 15 (2014) 100–112.
[5] Niko Kiukkonen, Jan Blom, Olivier Dousse, Daniel Gatica-Perez, Juha Laurila, Towards rich mobile phone datasets: lausanne data collection campaign, in: Proc. ICPS, Berlin, 2010.
[6] Phillipa Gill, Martin Arlitt, Zongpeng Li, Anirban Mahanti, Youtube traffic characterization: a view from the edge, in: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, ACM, 2007, pp. 15–28.
[7] Andreas Kaltenbrunner, Vicenç Gómez, Ayman Moghnieh, Rodrigo Meza, Josep Blat, Vicente López, Homogeneous temporal activity patterns in a large online communication space, 2007. ArXiv Preprint arXiv:0708.1579.
[8] Taha Yasseri, Robert Sumi, János Kertész, Circadian patterns of wikipedia editorial activity: a demographic analysis, PLoS One 7 (1) (2012) e30091.
[9] M. ten Thij, S. Bhulai, P. Kampstra, Circadian patterns in twitter, Data Anal. (2014) 12–17.
[10] Mehwish Nasim, Muhammad Usman Ilyas, Aimal Rextin, Nazish Nasim, On commenting behavior of facebook users, in: 24th ACM Conference on Hypertext and Social Media, ACM, 2013, pp. 179–183.
[11] Conrad Lee, Bobo Nick, Ulrik Brandes, Pádraig Cunningham, Link prediction with social vector clocks, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2013, pp. 784–792.
[12] Mehwish Nasim, Ulrik Brandes, Predicting network structure using unlabeled interaction information, in: MMB & DFT 2014 Proceedings of the International Workshop SOCNET 2014, University of Bamberg Press, 2014, pp. 57–64.
[13] Lionel Tabourier, Anne-Sophie Libert, Renaud Lambiotte, Predicting links in ego-networks using temporal information, EPJ Data Sci. 5 (1) (2016) 1–16.
[14] Nathan Eagle, Alex Sandy Pentland, David Lazer, Inferring friendship network structure by using mobile phone data, Proc. Natl. Acad. Sci. 106 (36) (2009) 15274–15278.
[15] Julián Candia, Marta C. González, Pu Wang, Timothy Schoenharl, Greg Madey, Albert-László Barabási, Uncovering individual and collective human dynamics from mobile phone records, J. Phys. A 41 (22) (2008) 224015.
[16] Zhi-Qiang Jiang, Wen-Jie Xie, Ming-Xia Li, Boris Podobnik, Wei-Xing Zhou, H. Eugene Stanley, Calling patterns in human communication dynamics, Proc. Natl. Acad. Sci. 110 (5) (2013) 1600–1605.
[17] Gautier Krings, Márton Karsai, Sebastian Bernhardsson, Vincent D. Blondel, Jari Saramäki, Effects of time window size and placement on the structure of an aggregated communication network, EPJ Data Sci. 1 (4) (2012) 1–16.

[18] Mehwish Nasim, Aimal Rextin, Numair Khan, Muhammad Muddassir Malik, On temporal regularity in social interactions: predicting mobile phone calls, 2015. ArXiv Preprint arXiv:1512.08061.
[19] Andres Sevtsuk, Carlo Ratti, Does urban mobility have a daily routine? learning from the aggregate data of mobile networks, J. Urban Technol. 17 (1) (2010) 41–60.
[20] Santi Phithakkitnukoon, Ram Dantu, Rob Claxton, Nathan Eagle, Behavior-based adaptive call predictor, ACM Trans. Auton. Adapt. Syst. (TAAS) 6 (3) (2011) 21.
[21] Osama O. Barzaiq, Seng W. Loke, Adapting the mobile phone for task efficiency: the case of predicting outgoing calls using frequency and regularity of historical calls, Pers. Ubiquitous Comput. 15 (8) (2011) 857–870.
[22] Mohamed Ramzi Haddad, Hajer Baazaoui Zghal, Djemel Ziou, Henda Ben Ghézala, A predictive model for recurrent consumption behavior: an application on phone calls, Knowl.-Based Syst. 64 (2014) 32–43.
[23] Albert-Laszlo Barabasi, The origin of bursts and heavy tails in human dynamics, Nature 435 (7039) (2005) 207–211.
[24] Hang-Hyun Jo, Márton Karsai, János Kertész, Kimmo Kaski, Circadian pattern and burstiness in mobile phone communication, New J. Phys. 14 (1) (2012) 013055.
[25] Ye Wu, Changsong Zhou, Jinghua Xiao, Jürgen Kurths, Hans Joachim Schellnhuber, Evidence for a bimodal distribution in human communication, Proc. Natl. Acad. Sci. 107 (44) (2010) 18803–18808.
[26] Hayang Kim, Hui Zang, Xiaoli Ma, Analyzing and modeling temporal patterns of human contacts in cellular networks, in: 2013 22nd International Conference on Computer Communications and Networks, (ICCCN), IEEE, 2013, pp. 1–7.
[27] Alessio Cardillo, Giovanni Petri, Vincenzo Nicosia, Roberta Sinatra, Jesús Gómez-Gardeñes, Vito Latora, Evolutionary dynamics of time-resolved social interactions, Phys. Rev. E 90 (5) (2014) 052825.
[28] Goolge Inc. Adaptive contact list, us 7797293 b2, 2010..
[29] Theophanis Tsandilas, et al., An empirical assessment of adaptation techniques, in: CHI'05 Extended Abstracts on Human Factors in Computing Systems, ACM, 2005, pp. 2009–2012.
[30] Frank Bentley, Ying-Yu Chen, The composition and use of modern mobile phonebooks, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, 2015, pp. 2749–2758.
[31] Antti Oulasvirta, Mika Raento, Sauli Tiitta, 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, 2005, pp. 167–174.
[32] Mehwish Nasim, Raphael Charbey, Christophe Prieur, Ulrik Brandes, Investigating investigating link inference in partially observable networks: friendship ties and interaction, 2016.
[33] T. Aledavood, E. López, S.G.B. Roberts, F. Reed-Tsochas, E. Moro, R.I.M. Dunbar, J. Saramäki, Daily rhythms in mobile telephone communication, ArXiv E-Prints arXiv:1502.06866 February 2015.
[34] T. Aledavood, E. López, S.G.B. Roberts, F. Reed-Tsochas, E. Moro, R.I.M. Dunbar, J. Saramäki, Channel-Specific daily patterns in mobile phone communication, July 2015. ArXiv E-Prints arXiv:1507.04596.
[35] Talayeh Aledavood, Sune Lehmann, Jari Saramäki, On the digital daily cycles of individuals, 2015. ArXiv Preprint arXiv:1507.08199.
[36] Eviatar Zerubavel, Hidden Rhythms: Schedules and Calendars in Social Life, Univ of California Press, 1985.
[37] Mehwish Nasim, Aimal Rextin, Numair Khan, Muhammad Muddassir Malik, Understanding call logs of smartphone users for making future calls, in: 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, ACM, 2016.
[38] Giovanna Miritello, Rubén Lara, Manuel Cebrian, Esteban Moro, Limited communication capacity unveils strategies for human interaction, Sci. Rep. 3 (2013).
[39] Ofer Bergman, Andreas Komninos, Dimitrios Liarokapis, James Clarke, You never call: demoting unused contacts on mobile phones using dmtr, Pers. Ubiquitous Comput. 16 (6) (2012) 757–766.
[40] Kenneth T. Rosen, Mitchel Resnick, The size distribution of cities: an examination of the pareto law and primacy, J. Urban Econ. 8 (2) (1980) 165–186.
[41] David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, Friedrich Leisch, e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, 2015. R package version 1.6-7.
[42] Anas Bilal, Aimal Rextin, Ahmad Kakakhel, Mehwish Nasim, Roman-txt: forms and functions of roman urdu texting, in: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, ACM, 2017.
[43] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008.