

An Empirical Analysis of the Human Dynamics Based on Mobile Phone Records

Haodong Lyu^{1, a}, Wenjun Wang¹, Pengfei Jiao¹ and Xiangyu Li¹

¹Tianjin Key Laboratory of Advanced Networking (TANK),
School of Computer Science and Technology, Tianjin University, Tianjin China, 300354

^ahlyu@tju.edu.cn

Keywords: Mobile phone records, human dynamics, inter-event time distribution, radius of gyration.

Abstract. Mobile phones are widely used by people in modern society, the records of which provide us with an essential proxy for the understanding of human behaviors. Previous empirical results have already shown that some temporal, even spatial, characteristics of many human activities follow power law instead of Poisson statistics. In this paper, we investigate the statistics of human mobile phone communication patterns and their mobility patterns based on mobile phone records. The results show that the distributions of related person and the distributions of inter-event time for four different behaviors all follow power law distribution with different values of exponent. In addition, we have a comparative analysis of human behavior in four different periods of a day, and find out that most of individuals are more active in the afternoon and less active at night in aspects of the diversities of persons they reached and the range of their activities.

Introduction

Mobile phones, due to their characteristics of portability, convenience and flexibility, are widely used in modern society. The uses of mobile phones, such as making phone call, answering the phone, sending and receiving text messages, inevitably leave some digital traces. Moreover, even if you are sitting there and doing nothing with your phone, it will upload some information to the server periodically. Thus, the extremely valuable records of mobile phone provide us with great opportunities to study human mobility patterns, communication dynamics and social structure [1].

Jukka-Pekka Onnela et al. construct a large-scale calling network from European data and analyze many basic and advanced network characteristics [2]. Ming-Xia Li et al. investigate the statistical properties of four calling networks (DCN, MCN, SVDCN, and SVMCN) which are constructed from the data of more than 9 million individuals in Shanghai over 110 days [1]. There is no doubt that it is very helpful for the study of social structure from the perspective of complex networks using massive mobile phone communication data.

The studies on the temporal pattern of human behavior are also elaborated [3, 4]. After analyzing the inter-event time distribution for calling activity of six million mobile phone users during a month, Julián Candia et al. find out that the data can be well fitted by power law with an exponential cutoff [5]. ZHAO Zhi-Dan et al., using the SM communication records from 27 to 29 January 2006, find out that the distribution of SM sending, the inter-event time distribution of SM sending and distribution of response time are all heavy-tailed and they can all be partially fitted by power law [6]. These communication patterns are also observed in other situations such as network communications [7, 8], financial activities [9, 10, 11], web browsing [12, 13, 14, 15] and so on.

Marta C. González's team may be one of the earliest teams who concentrated on the study of human mobility patterns using mobile phone data [16]. In the year of 2012, After analyzing large population movements caused by Haiti earthquake, Xin Lu et al believe that their findings suggest that population movements can be predictable during disasters [17].

In this paper, to analyze the temporal patterns of human communication behaviors and their mobility patterns based on mobile phone records data, we study the number of related person distribution, inter-event time distribution and radius of gyration distribution. We also have a comparative analysis of different periods of a day including night, morning, afternoon and evening period for all the

measurements mentioned above, on which no research has been carried out yet according to our knowledge.

Dataset and Methods

Dataset Description. The dataset we used contains the detailed mobile phone records of more than 36 million anonymous mobile users from one operator in a metropolitan city of China. This sample has a total of over three weeks (24 days) mobile phone records from Oct 8, 2015 to Oct 31, 2015. The following columns are included by the structured records, event ID, starting time of the event, base station ID and its latitude, longitude coordinates, caller number, called number and other extra information. Both the caller number and the called number have been encrypted so that there are no privacy issues. By different types of event, the sample is categorized into six parts for convenience. In order to avoid interference caused by various noise, only the records of active mobile phone users are used. In term of a specific event, active mobile phone users are defined as someone whose records show up at least once every day in this period. The detailed information about the data is showed in table 1. Please notice that called numbers of the last two events are always null in each record.

Table 1 Detailed information about data

Event type	Making phone call	Answering the phone	Sending text message	Receiving text Message	Mobility location updating	Periodic updating
Number of users	159,274	109,684	8,106	3,884	220,084	155,484
Number of records	43660346	27009164	2456173	171202122	81916472	36943407

Methods

Inter-event Time. Inter-event time τ is defined as the interval between two same and consecutive events. In our case, it includes inter-event time of making phone calls, answering the phone, sending text messages and receiving text messages respectively. To be specific, $\tau = t_i - t_{i-1}$, where t_{i-1} is the exact time when the $(i - 1)_{th}$ event happened and t_i is the exact time when the i_{th} event happened. The probability distribution is defined as $P(\tau)$.

Radius of Gyration. Assume $T_i = \{t_1, t_2, \dots, t_{L_i}\}$ as the sequence of mobile phone base stations that the phone number of individual i has been recorded during a period and $\mathbf{r}(t)$ as the location of t . Then the radius of gyration of i 's trajectory in this specific period can be defined as

$$r_g(i) = \sqrt{\frac{1}{L_i} \sum_{k=1}^{L_i} (\mathbf{r}(t_k) - \bar{\mathbf{r}})^2}, \quad (1)$$

where $\bar{\mathbf{r}} = \frac{1}{L_i} \sum_{k=1}^{L_i} \mathbf{r}(t_k)$ is the center of mass of the trajectory [17].

The radius of gyration is different from the daily average travel distance. Even if someone moves a long distance in a comparatively confined space, he or she will have a relative small value of radius gyration.

Four Periods of a Day. To study different patterns of human behaviors over a day, we divide in into four periods, namely night: 0 am – 6 am, morning: 6 am – 12 pm, afternoon: 12 pm – 6 pm, evening: 6 pm – 0 am (the next day). Since the interval time exerts serious effects on the results, it is essential to divide a day into four equal parts.

Results and Discussion

Number of Related Person Distribution. For a specific individual, the number of related person means that the number of different persons involved by one of his/her behaviors. For instance, in term of one person's behavior of making phone calls, the number of related persons equals the number of all persons he or she called. As we are aiming to find out the variant patterns of different behaviors, the first thing we should know is the number of persons contacted by each individual in a day. Thus, a general comparison of the distributions of the four events is shown in Fig. 1.

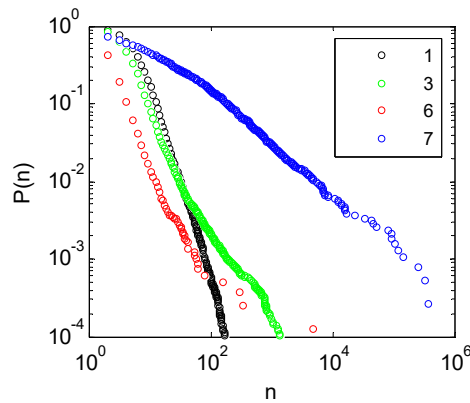


Fig. 1 Distribution of the number of related person for different types of events. The number “1”, “3”, “6”, “7” in the legend represent the behavior of making phone call, answering phone call, sending text message and receiving text message respectively.

As we can see, these four curves are skewed with a fat tail, which indicate that most mobile phone users communicate with a few individuals while a small minority interacts with dozens. And they all can be fitted or partially fitted by a power law function. A quite remarkable difference between receiving text message and the other three behaviors is that its exponent (slope in double logarithmic coordinates) is much smaller than others and its range is much wider. The exponents for different types of event are 3.46, 2.95, 2.94, and 1.70 respectively. It is a quite reasonable phenomenon. Readers may check the text message senders in your mobile phone message boxes if the number of senders is much larger than the number of persons in your contacts. Due to the wide use of SMS reminders, all the websites you signed up with your mobile phone number may send you a text message when you have certain behaviors in their websites (purchase stuff on an online shopping site, for instance). Also the curve of sending text message (red) is lower than making phone call (black), which demonstrates that making phone call is more popular than sending text message among interpersonal communications. The irregular shape of the distribution for answering the phone caused by the fatter tail in comparison with the behavior of making phone call has been observed in [2]. In particular, while the most connected user in making phone call network has the $k_{max} = 517$, in the answering phone call network we find that $k_{max} = 6903$.

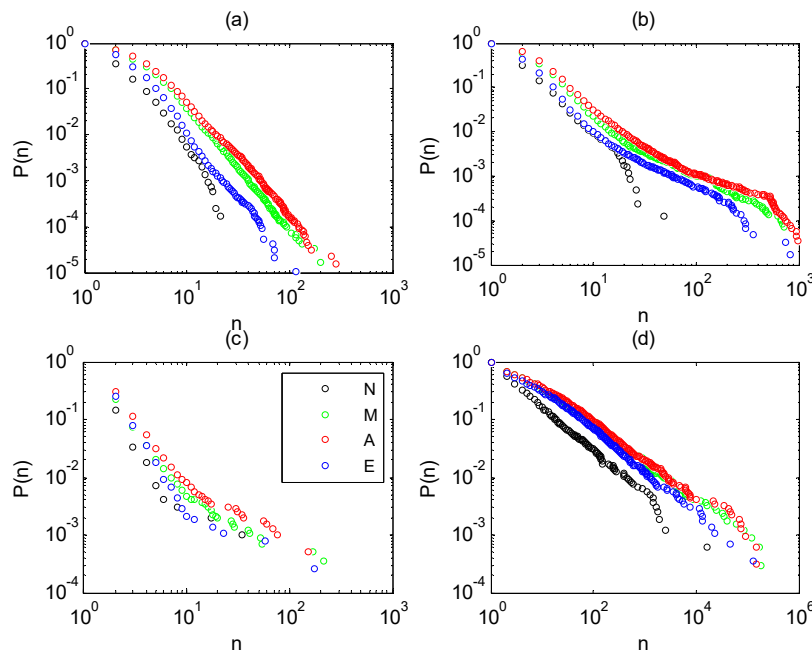


Fig. 2 Distribution of the number of related person for different periods of a day. (a) Making phone call; (b) answering the phone; (c) sending text message; (d) receiving text message. The letter “N”, “M”, “A”, “E” in the legend are beginning of “Night”, “Morning”, “Afternoon” and “Evening” respectively.

Fig. 2 shows the comparison of four different periods within a day for different types of event respectively. It can be seen that all the black curves (represent night) in every sub-figure have the narrower range and are lower than others, followed by the curves of the evening and morning curves. The curve of afternoon is the highest among the four and has the widest range, which demonstrates that most individuals have the highest activity level in the afternoon and the people they connected are more diversified, followed by morning and evening curves. Most people behave more quiet at night. It is congenial with reason and common sense.

Inter-event Time Distribution. Fig. 3 illustrates the inter-event distribution of four different types of events respectively (time resolution, seconds). The four curves have similar shape as Fig. 2. in [6]. When the interval time is less than 30 seconds approximately, the curves fluctuate. In the range $\tau \in [30, 20000]$, they all look like a straight line. And there is a cutoff followed at $\tau \approx 20000$ respectively. We believe that the very short interval time occurred in the behavior of sending text message caused by mass message sending, i.e., a mobile phone user will send a same message to a list of contacts after the message is written. Besides, the fitted line of inter-event distribution of receiving text message has the largest slope value, and is the lowest among the four, the phenomenon uncovered by which is consistent with what is shown in Fig. 1. People in modern society not only receive a large amount of text messages from various sources, but also receive them more frequently compared with the behavior of answering mobile phone. The position of the red curve demonstrates that the time intervals of users' sending text message is narrower than making phone calls even if they are accustomed to communicating with making phone calls. The distributions of answering the phone and making phone calls are so similar that the two curves nearly coincide. However, based on the exponent value from table 2, the difference still exists.

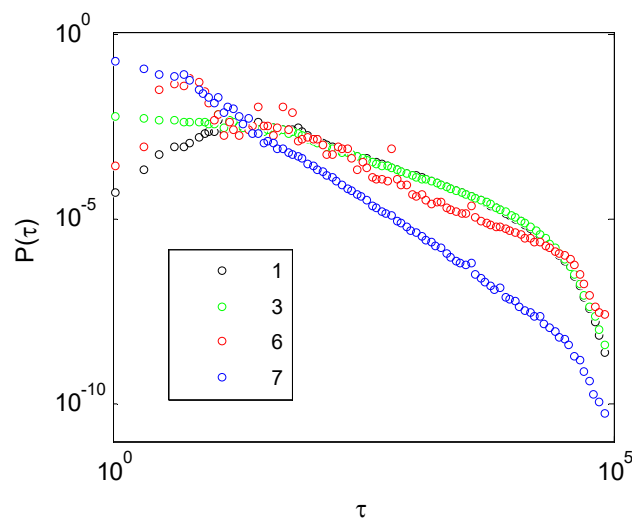


Fig. 3 Distribution of inter-event time for different types of events. The number “1”, “3”, “6”, “7” in the legend represent the behavior of making phone call, answering phone call, sending text message and receiving text message respectively.

Table 2 Exponents of fitted power law (fitted by the data in range [30,20000])

Event type	Making phone call (1)	Answering the phone (3)	Sending text message (6)	Receiving text Message (7)
Exponent	1.085	0.9774	1.235	1.682

Fig. 4 shows the comparison of the inter-event time distribution of four different period within a day for different types of events respectively (time resolution, seconds). From Fig.4 (d) we can see that the curve of morning has the smallest slope, followed by evening, afternoon and morning, which means that the frequency of receiving text message decreases at night. Fig. 4(c) illustrates a counterintuitive phenomenon that the interval time of most users' sending text messages at night are much shorter than other periods within a day. This phenomenon is explainable in that when night falls, you are lying in bed and chatting with friends using text messages. Without being disturbed by

various factors you may encountered at daytime, you are anxiously waiting friends' reply and will response immediately as soon as their message arrives. The four curves in the other two subfigures are so close that it seems that they do not have difference at all. It uncovers that even the diversities of people they connected in different periods are not different, but the interval times of making phone call have little difference in the four different periods. This same phenomenon holds for answering phone calls.

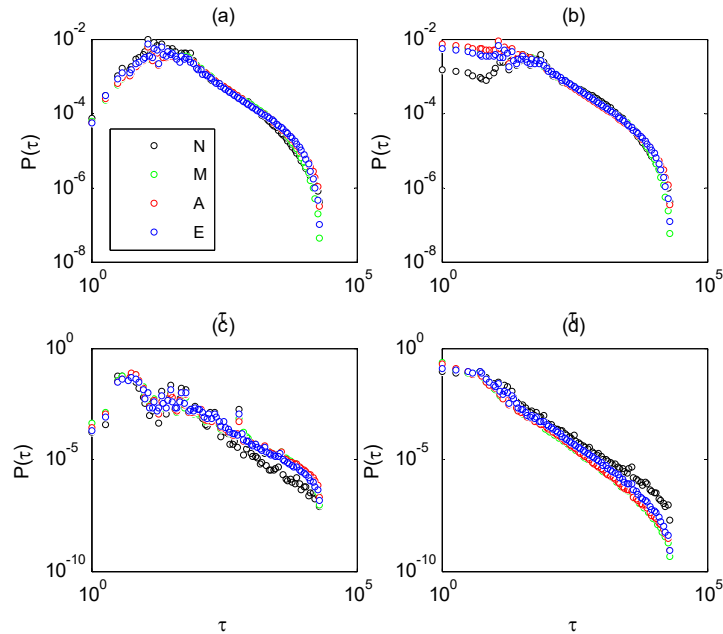


Fig. 4 Distribution of inter-event time for different periods of a day. (a) Making phone call; (b) answering the phone; (c) sending text message; (d) receiving text message. The letter “N”, “M”, “A”, “E” in the legend are beginning of “Night”, “Morning”, “Afternoon” and “Evening” respectively.

Radius of Gyration Distribution. To analyze people’s trajectories in different periods within a day, we focus on the radius of gyration, a measure of the size of trajectories. To prepare the calculation of radius of gyration, the digital footprint of each person is needed. So we mix all six different types of events up. More information about radius of gyration can be seen in the Methods section.

From Fig.5 (a) we can see the distribution of individuals’ radius of gyration within a day. At the beginning, the curve fluctuates, followed an increase. In the range $R_g \in [300m, 50000m]$, it has a relatively stable distribution and high probability, which indicates that radiuses of gyration of the majority are in this range. Besides, 300m is the coverage of mobile phone base station. With the value of R_g increases continuously, there is a rapid decrease. Notice that when $R_g \approx 10km$, a small peak turns up. The fact is that radius of central area of this city is ten kilometers approximately.

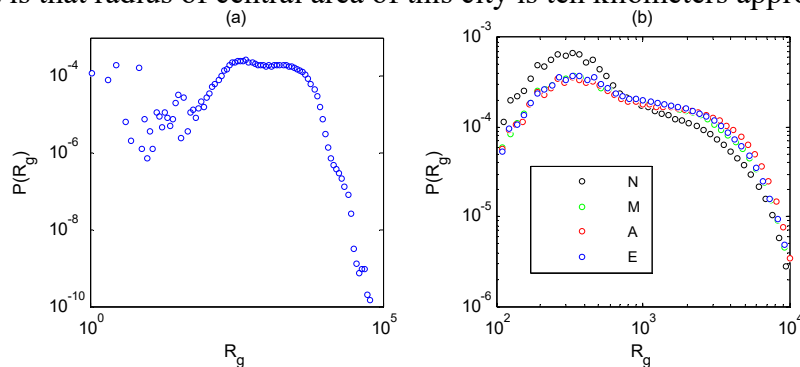


Fig. 5 (a) Distribution of radius of gyration in a day; (b) Distribution of radius of gyration for different periods of a day. The letter “N”, “M”, “A”, “E” in the legend are beginning of “Night”, “Morning”, “Afternoon” and “Evening” respectively.

Fig. 5(b) shows the distribution of different periods of a day, in which only the curves within the range between 100m and 10km are displayed. It is obvious that at night, people’s radiuses of gyration

are relatively small, when people may be walking around their houses. Moreover, In the afternoon, people seem to have a broader range of activities compared to the other three periods within a day.

Conclusions

Research on human behavior has been attracting researchers' attention. Along with them, we have empirically studied some temporal and spatial properties of human's behaviors in this paper, including four different types of event. Besides, we have made a comparison among four different periods of a day, i.e., the night, morning, afternoon and evening period. Both the number of related person and inter-event time follow power law distribution with a heavy tail, which means that most mobile phone users are less active yet a minority of users are very active. In term of diversities of people are reached, making phone call is more popular than sending text message among interpersonal communications. And people are receiving more messages from a variety of sources. Due to the distinct function of phone and message, the inter-event time of most individuals' sending or receiving message are shorter than making phone call or answering the phone. By comparing the related person distribution and radius of gyration distribution of the four periods of a day, we find out that individuals are more active in the afternoon and less active at night. It's hard to tell the difference among the four periods of a day in term of interval time of making phone call or answering phone call, but it is obvious that the interval time of sending message is more shorter at night while the interval time of receiving message is more shorter at daytime. The reason is that sending message is a more active behavior while receiving is more passive to some extent. Moreover, we find that the ranges of activity of most individuals are between three hundred meters and five kilometers. All these findings not only shed light on collective human behaviors, but also provide a probe for further research.

In the future, in order to reveal deeper mechanisms of human dynamics, we will continue to analyze the correlation between temporal pattern and spatial pattern. To achieve that, a large amount of work is awaiting us, which we believe is extremely meaningful and interesting.

Acknowledgement

This work was supported by the Major Project of National Social Science Fund (14ZDB153), the major research plan of the National Natural Science Foundation (91224009, 51438009), the National Science and Technology Pillar Program (2015BAL05B00 and 2015BAL05B02).

References

- [1] Li, M. X., Jiang, Z. Q., Xie, W. J., Miccichè, S., Tumminello, M., & Zhou, W. X., et al. (2014). A comparative analysis of the statistical properties of large mobile phone calling networks. *Scientific Reports*, 4(3), 5132-5132.
- [2] Onnela, J. P., Saramäki, J., Hyvönen, J., Szabó, G., Menezes, M. A. D., & Kaski, K., et al. (2007). Analysis of a large-scale weighted network of one-to-one human communication. *New Journal of Physics*, 9(26).
- [3] Hong, W., Han, X., Zhou, T., & Wang, B. (2008). Heavy-tailed statistics in short-message communication. *Chinese Physics Letters*, 26(2), 297-299.
- [4] Wu, Y., & Schellnhuber, H. J. (2010). Evidence for a bimodal distribution in human communication. *Proceedings of the National Academy of Sciences*, 107(44), 18803-8.
- [5] Candia, Julián, González, Marta, C., Wang, & Schoenharl, et al. (2007). Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A Mathematical & Theoretical*, 41(22), 1441-1446.
- [6] Zhao, Z. D., Hu, X., Shang, M. S., & Tao, Z. (2011). Empirical analysis on the human dynamics of a large-scale short message communication system. *Chinese Physics Letters*, 28(6), 68901-68904(4).

- [7] Radicchi, F. (2009). Human activity in the web. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 80(2), 1711-1715.
- [8] Leskovec, J., & Horvitz, E. (2010). Planetary-Scale Views on a Large Instant-Messaging Network. (pp.915-924). ACM.
- [9] Vázquez, A., Oliveira, J. G., Dezsö, Z., Goh, K. I., Kondor, I., & Barabási, A. L. (2006). Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73(3 Pt 2), 80-98.
- [10] Scalas, E., Kirchler, M., Kaizoji, T., Huber, J., & Tedeschi, A. (2006). Waiting times between orders and trades in double-auction markets. *Physica A Statistical Mechanics & Its Applications*, 366(C), 463–471.
- [11] Wang, S. C., Tseng, J. J., Tai, C. C., Lai, K. H., Wu, W. S., & Chen, S. H., et al. (2007). Network topology of an experimental futures exchange. *Physics of Condensed Matter*, 62(1), 105-111.
- [12] Dezsö Z, Almaas E, Lukács A, Rácz B, Szakadát I, & Barabási AL. (2006). Dynamics of information access on the web. *Physical Review E*, 73(6 Pt 2), págs. 60-61.
- [13] Gonçalves, B., & Ramasco, J. J. (2008). Human dynamics revealed through web analytics. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 78(2 Pt 2), 026123-026123.
- [14] Zhao, G., Zhang, N., Liu, Z., & Li, J. (2010). A Case Investigation on the Scaling Behaviors in Web Browsing. *Ieee/wic/acm International Conference on Web Intelligence and Intelligent Agent Technology (Vol.3, pp.160-163)*. IEEE Computer Society.
- [15] Chun, H., Ahn, Y. Y., Kwak, H., Moon, S., Eom, Y. H., & Jeong, H. (1970). Comparison of online social relations in terms of volume vs. interaction. *Annals of Physics*, 61(2), 488-529.
- [16] González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779-82.
- [17] Lu, X., Bengtsson, L., & Holme, P. (2012). Predictability of population displacement after the 2010 haiti earthquake. *Proceedings of the National Academy of Sciences of the United States of America*, 109(29), 11576-11581.