User Contact Behavior Analysis: A Social Relationship Perspective

Peiyan Yuan, Hai Yu, Ping Liu
{peiyan@htu.cn, yuhaihtu@foxmail.com, lp@htu.cn}

School of Computer and Information Engineering, Henan Normal University
46# East Construction Road Xinxiang, Henan, China 453007

Abstract. Mobile opportunistic networks (MONs) have been attracting increasing amounts of attention in recent years. Characterizing user contact behavior provides a baseline to evaluate the performance of these networks. However, because the contact distribution of nodes in MONs is conventionally modeled from a large-scale perspective, i.e., by aggregating all node pairs, the contact features of nodes with multiple social relationships are not reflected. Thus, it is not clear whether friends and strangers have similar or different contact behaviors. In this study, we aggregated the contact information of users from the real world, and discovered that two phenomena exist: (1) Most friends or strangers make contact at public hotspots, rather than private hotspots; (2) The distribution of intra-contact time (ICT) exhibits different decay factors—the ICT distribution of strangers is predominantly faster than that of friends.

Keywords: Mobile opportunistic networks; User contact behavior; Social relationship; Hotspots.

1 Introduction

Concomitant with the significant advancements in microelectronic and communication technologies, numerous sensors and wireless interfaces are now integrated into portable devices such as smart-phones and personal digital assistants (PDAs). This has resulted in the emergence of a new communication paradigm called Mobile Opportunity Networks (MONs). In contrast to traditional mobile cell networks and ad hoc networks, MONs do not have a connected path between sources and destinations. Instead, messages are transmitted with a store-carry-and-forward style. This new feature in MONs creates various exciting opportunities and many interesting applications, ranging from mobile offloading [1] to urban computing [2], are being envisioned.

Characterizing user contact behavior provides a foundation to evaluate the performance of MONs, because the mobility of people plays a significant role in the applications being envisioned [3] [4] [5]. For example, before supplementary municipal planning facilities can be deployed in Beijing, a key issue would be determination of the potential. Collecting user's trajectories and detecting their activities in order to ascertain the primary gathering places may be a good choice [6].

Over the past years, several user mobility analysis studies have been conducted [7] [8], where two primary contact features are considered, inter-contact time and intra-contact time [1] [9][10] [11] [12] . Numerous experiments have revealed that both features follow a power-law

[3] [13] or an exponential distribution [5] [14] [15]. However, because most studies model the contact distribution of nodes from a large-scale perspective, i.e., by aggregating all node pairs, the contact features of nodes with multiple social relationships is not reflected [13]. For example, their models cannot determine whether friends and strangers have similar or different contact behaviors.

Taking this issue into account, we characterize user contact behavior from a social relationship perspective [14] [16] [17]. Specifically, we first divide the Access Points (APs) into two categories---public hotspots (i.e., APs with higher weights) and private hotspots (i.e., APs with lower weights)---and classify users into friends and strangers, respectively. We call two users friends if their contact duration is greater than the average contact time; otherwise, we call them strangers [18]. Analysis of their GPS traces revealed the existence of two phenomena: (1) Most friends or strangers make contact at public hotspots, instead of private hotspots; (2) The distribution of intra-contact time (ICT) exhibits different decay factors—the ICT distribution of strangers is predominantly faster than that of friends.

The remainder of this paper is organized as follows. We first discuss the related work and introduce preliminaries. Next, we outline our proposed classification method for hotspots and user's relationship. Subsequently, we present and analyze the results experiments conducted. Finally, we present concluding remarks.

2   Related work

Social interaction plays an important role in human community applications. An accurate and reasonable model is vital to reflect the spread of a disease, characterize the traffic of urban transportation etc. For example, the authors of [19] focused on the impact of the epidemic on the social networks. Sekara et al. [20] analyzed the interaction between the mobile social network and the physical space.

Much research, therefore, has been conducted on the distribution of inter-contact time and intra-contact time in mobile opportunistic networks [13]. For example, Mcnett and Voelker [21] studied the access behaviors of wireless PDA users by first collecting a trace from 275 freshmen with HP Jornada PDAs over 11 weeks, from September 22, 2002 to December 8, 2002. They developed a tool called the wireless topology discovery (WTD) tool to periodically collect information on each PDA installation and studied user's mobility and analyzed the number and duration of AP sessions. They also developed two wireless network topology models and used the evolutionary network topology model as a case study to evaluate ad hoc routing algorithms.

Natarajan and Motani et al. [11] subsequently analyzed user interactions from Bluetooth contact traces. In their study, they collected more than 350,000 contact records from 10,000 unique devices, and proposed several new metrics to evaluate user behaviors. They found that most of their metrics followed a power law, except for inter-pair-contact time. The predictability of these metrics facilitates the design of smart routing algorithms. Chaintreau et al. [9] also studied the impact of human mobility on opportunistic forwarding schemes. They employed a simplified model that incorporated the renewal theory, and proved that the power-law condition can be solved with a finite expected delay by oblivious forwarding algorithms, but only if the heavy tailed index of the distribution is greater than one. That is to say, when the heavy tailed index is less than one, it cannot limit the expected delay for any forwarding algorithm; the queue in the relay device is ignored.

Similarly, Karagiannis et al. [10] studied the distribution of inter-contact times between mobile devices and subsequently proposed the dichotomy hypothesis, which states that the distribution of intercontact time obeys the power law until it has decayed to a certain point, beyond which it becomes an exponential distribution. Yuan et al. [22] studied human behavior in both virtual and physical spaces, using inter-action time (IAT) to reflect the relationship of people in both spaces. They subsequently stated that the complementary cumulative distribution function (CCDF) of IAT also follows a power law with an exponential decay.

Moon and Helmy [15] investigated the periodicity and regularity of nodal encounter using the power spectral. They found that mobile nodes with few encounters have strong periodicity, and vice versa. Wei et al. [23] studied the specific movement characteristics of nodes while taking social relationships into account. They divided social ties into two types: static and dynamic. They considered static social relations to have long time granularity and stabilize over time, whereas dynamic social relations have relatively short time granularity and may be change over time. They found that a node pair from the same community makes contact more frequently than those from different communities. Therefore, they classified node pairs into two classes: familiar and unfamiliar. Their experimental results showed that familiar pairs and unfamiliar pairs of nodes have the same inter-contact time distribution, and that both approximately followed the power-law decay to a certain point, beyond they are subject to exponential decay. Their results also indicated that the inter-contact time distribution of familiar pairs decays faster than inter-contact time distribution of unfamiliar pairs. In other words, the probability of contact for familiar node pairs is much greater than that for unfamiliar node pairs.

3   Preliminaries

We used the well-known Dartmouth datasets [24] to analyze the distribution of user contacts. The original data was collected with 20 Nokia N95 phones carried by students and staff members at Dartmouth College from July 28 to August 11, 2008. The dataset includes the following information for each user: accelerometer raw data and GPS location coordinates. Considering the same periodicity of human mobility, we extract data pertaining to a period of seven days, from August 2 to 8, 2008. The data comprised 1838 users, 129 hotspots, and 6119 records.

Preprocessing: We first converted the hexadecimal data to decimal and used a five-tuples set (User ID, AP ID, Turn-on time, Turn-off time, Contact time) to denote the contact between a user and an AP, where start time denotes the moment when a user enters the coverage range of an AP and end time denotes the moment when the link between them is disconnected. The difference between the end time and the start time is the contact duration. Table 1 examples of the various parameter values in the dataset.

Table 1. Part of records of user contacts.

User ID	AP ID	Turn-on	Turn-off	Contact time
…	…	…	…	…
1	104	28832	31534	2700
1	104	81292	81302	10
1	104	84419	84429	10
1	104	28465	31449	3000
1	104	40395	44482	4080
1	104	44966	50372	5400
1	104	66383	71268	4860
1	104	75202	75212	10
1	104	78538	80930	2340
…	…	…	…	…
2	88	83515	62	2940
2	88	80343	86288	5940
2	88	77503	81905	4440
2	88	81905	81915	10
3	0	77173	79505	2340
4	0	71826	75578	3720
4	0	48207	48858	660
…	…	…	…	…

3.1   Hotspot Weights

The weights of hotspots characterize their popularities. In general, the larger the value is, the more popular is that hotspot. We here take a simple but efficient solution to estimate the weight of hotspots, called count process[25]. This method calculates the number of stay points in each hotspot and normalizes the sampling counts, so as to calculate the weight of each hotspot. Let c denote the number of total contacts between users and APs within a week, and ci represent the contact number between users and the ith hotspot, we have

$$w_i = c_i / c \quad (i = 1, 2, 3...)$$ (1)

Where w_i denoted the weight of the i_{th} hotspot.

Using Equation 1, we plotted the weight distribution of all the hotspots, as shown in Figure 1(a). From the figure, it is clear that the weights of hotspots fluctuate considerably. The maximum fluctuation is over 5%, whereas the minimum tends to zero. In order to classify hotspots, we sort the hotspots according to the weight, as shown in Figure 1(b). Subsequently, we got the median of 128 nodes, 0.005, as a threshold, and obtained 64 private hotspots (with

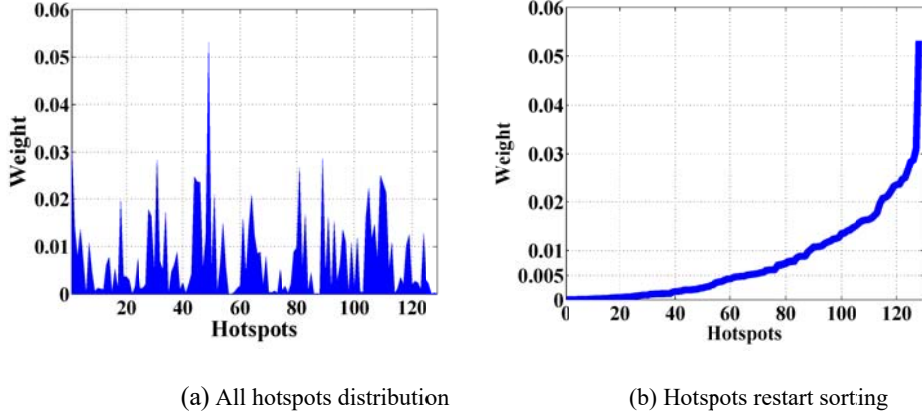lower   weight   AP),   and   the   rest   are   public   hotspots.

(a) All hotspots distribution   (b) Hotspots restart sorting

Fig. 1. Hotspots weight distribution.

3.2   Identifying Social Relationship

In this paper, a contact is defined as two users simultaneously accessing the same AP, and the intra contact time is defined as the duration of time both of them are connected to the same AP. As in [23], we use the average intra-contact time to identify a user's social relationships. Mathematically, let $x_i(j)$ denote the intra-contact time between user i and user j, and let random variable X_i denote the intra-contact time between user i and other users except the user i. $E(X_i)$ represents the expectation of X_i. Then, we have

$$E\left(X_i\right) = \frac{\sum_{i,j\in N, i\neq j} x_i\left(j\right)}{\|N\|} \tag{2}$$

Where, the N denotes the set of users. When $xi(j)$ is bigger than $E(Xi)$, we call user j is a friend of user i; Otherwise, the user i and the user j are strangers.

4   Results

In this study, we classified user pairs into friends/strangers and analyzed their contact behaviors in public hotspots and private hotspots, respectively. We examined their contact behavior in the following four situations: (1) friends in private hotspots, (2) in public hotspots, (3) strangers in private hotspots, and (4) strangers in public hotspots.

4.1   Contact proportion in different hotspots

Table 2 shows the number of contacts by friends and strangers in both public and private hotspots. The table shows that both friends and strangers tend to make contact in public areas, rather than private areas.

Table 2. Proportion of friends and strangers making contact in public and private hotspots.

	Public hotspots	Private hotspots
Friends	97.1%	2.7%
Strangers	94.6%	5.4%

4.2 Analysis of contact distribution

This subsection focuses on the contact distribution of users. We first examine the contact distribution by integrating all users, and then analyze the difference in the ICT distribution between friends and strangers. Finally, we analyze the influence of the distribution of ICT.

4.2.1 Contact distribution of all user pairs

We collected the intra-contact time information of all user pairs in both public and private hotspots and plotted their corresponding CCDF, as shown in Figure 2. We consider two users as making contact when they are both in the same AP range. Obviously, the CCDF of ICT follows a heavy tail index, that is, the longer the ICT is, the smaller the contact time of the two users. We found that 3615s is a critical point as most of the ICTs are smaller than this value.
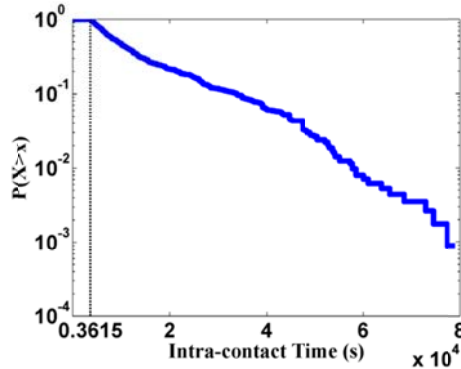

Fig. 2. Intra-contact time CCDF for all users.

4.2.2 Influence of social relationship on the distribution of ICT

This subsection focuses on the contact distribution of friends and strangers. In the above subsection, we showed that the ICT of all user pairs followes heavy tail index. Thus, it can be expected that the ICT of friends/strangers will also give similar results. In this case, however, we wish to determine the difference in the ICT between the two social relationships. In this process, we adopted the method developed by Chaintreau et al. [9] to estimates the parameter of the power law, also called the heavy tailed index. The approach is used to study the heterogeneity among user pairs.

Estimator for the Heavy Tail Index: Let $\{m_1, m_2, ..., m_n\}$ denote a set of sequences of the length of intra-contact time between all users, and m is the average of this sequences. Let T denote the time granularity. According to the CCDF of ICT, we have

$$P[X > x] = 1 - F[x] \approx cx^{-a} \tag{3}$$

since the value of T is 300s for the used datasets, we have

$$P[X > x] = (x / 300)^{-a} \tag{4}$$

where the heavy tail index/alpha equals to

$$\alpha = \frac{\ln 2}{\ln(m) - \ln(T)} \tag{5}$$

Figure 3 shows the CCDF of ICT for both friends and strangers. It is clear that even though both of them also follow heavy tail index, the CCDF of strangers decays faster than that of friends. Using Equation 5, we can obtain the value of α in both situations. The former is 0.2154 and the latter is 0.1508. The larger the value of α is, the faster the decay should be.
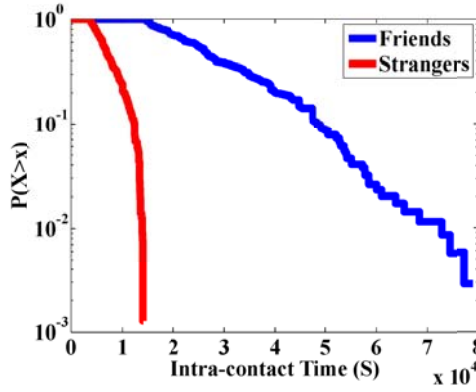

Fig. 3. Intra-contact time CCDF for friends and strangers.

4.2.3 Influence of hotspots on the distribution of ICT

We further analyzed the impact of hotspots on the ICT distribution. Figure 4(a) shows the CCDF of ICT for the contact time distribution of friends in private areas and in public areas, respectively. It can be seen that the ICT distribution of friends in private areas decays faster than that of friends in public areas. From Equation 5, we obtain the value of α for friends in public hotspots as 0.1507 and the value in private hotspots as 0.1615. We also found that when the user contact time exceeds 33958 s, the possibility of users making contact in private areas is virtually zero. This implies that people prefer to spend more time in public areas than that in private areas. Similarly, Figure 4(b) shows the CCDF of ICT for the contact time distribution of strangers in private areas and public areas, respectively. There is virtually no difference between the two situations: the value of α in public hotspots is 0.2150 and that in private hotspots is 0.2217, which implies that stranger contact time does not depend on the place where the encounter occurs.
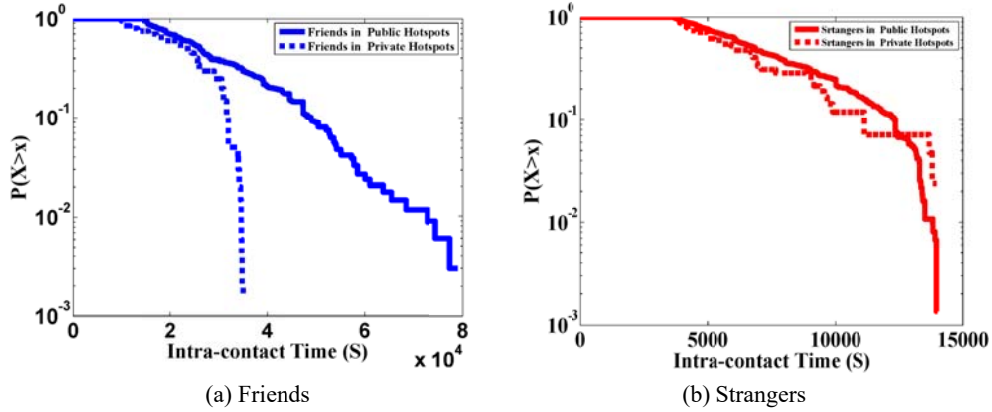
(a) Friends   (b) Strangers

Fig. 4. Intra-contact time CCDF for friends and strangers.

5   Conclusion

In this study, we investigated user contact behavior from the perspective of social relationship. We first grouped APs into public hotspots and private hotspots and the social relationship of users into friends and strangers. Then, we discovered that the distribution of intra-contact time closely follows the heavy tail index, and that both friends and strangers frequently make contact in public hotspots rather than private hotspots. Compared to strangers, friends have a faster decay in private areas, which implies that friends prefer to spend more time in public areas. In addition, the contact time of strangers was found to be independent of where they made contact. We believe that the results of this study will provide new insights into user contact behavior in mobile social networks.

References

[1] Huadong Ma, Dong Zhao, and Peiyan Yuan. Opportunities in mobile crowd sensing. IEEE Communications Magazine, 2014, 52(8):29-35.

[2] Lina Al-Kanj, H. Vincent Poor, and Zaher Dawy. Optimal cellular offloading via device-to-device communication networks with fairness constraints. IEEE Transactions on Wireless Communications, 2014, 13(8):4628-4643.

[3] Jie Bao, Yu Zheng, David Wilkie, and Mohamed Mokbel. Recommendations in location-based social networks: A survey. Geoinformatica, 2015, 19(3):525-565.

[4] John Krumm and Eric Horvitz. Locadio: Inferring motion and location from wi-fi signal strengths. In International Conference on Mobile and Ubiquitous Systems: NETWORKING and Services, 2004. MOBIQUITOUS, 2004, 4-13.

[5] Wan Shiou Yang, Hung Chi Cheng, and Jia Ben Dia. A location-aware recommender system for mobile shopping environments. Expert Systems with Applications, 2008, 34(1):437-445.

[6] Mauro Brunato and Roberto Battiti. A locationdependent recommender system for the web. Proceedings of the Mobea Workshop, 2002, (4):175-179.

[7] Ellen Spertus, Mehran Sahami, and Orkut Buyukkokten. Evaluating similarity measures: A large-scale study in the orkut social network. In Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, Illinois, USA, August, 2005, 678-684.

[8] Peiyan Yuan and Huadong Ma. Opportunistic forwarding with hotspot entropy. In World of Wireless, Mobile and Multimedia Networks, 2013, 1-9.

[9] Augustin Chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, and James Scott. Impact of human mobility on opportunistic forwarding algorithms. IEEE Transactions on Mobile Computing, 2007, 6(6):606-620.

[10] Thomas Karagiannis, Jean Yves Le Boudec, and Milan Vojnovic. Power law and exponential decay of intercontact times between mobile devices. ACM International Conference on Mobile Computing and NETWORKING. ACM, 2007:183-194

[11] Anirudh Natarajan, Mehul Motani, and Vikram Srinivasan. Understanding urban interactions from bluetooth phone contact traces. In Proceedings of Passive and Active Network Measurement, International Conference, PAM 2007, Louvain-LaNeuve, Belgium, April 5-6, 2007, 115-124.

[12] Andrea Passarella and Marco Conti. Characterising Aggregate Inter-contact Times in Heterogeneous Opportunistic Networks. NETWORKING 2011. Springer Berlin Heidelberg, 2011:301-313.

[13] Andrea Passarella and Marco Conti. Analysis of individual pair and aggregate intercontact times in heterogeneous opportunistic networks. IEEE Transactions on Mobile Computing, 2013, 12(12):2483-2495.

[14] Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, and Wei Ying Ma. Mining user similarity based on location history. In ACM Sigspatial International Symposium on Advances in Geographic Information Systems, Acm-Gis 2008, November 5-7, 2008, Irvine, California, USA, 2008, 1-10.

[15] Sungwook Moon and Ahmed Helmy. Understanding periodicity and regularity of nodal encounters in mobile networks: A spectral analysis. In Global Telecommunications Conference (GLOBECOM 2010), 2010, 1-5.

[16] Haifeng Liu, Zheng Hu, Ahmad Mian, Hui Tian, and Xuzhen Zhu. A new user similarity model to improve the accuracy of collaborative filtering. Knowledge-Based Systems, 2014, 56(3):156-166.

[17] Matteo Manca, Ludovico Boratto, and Salvatore Carta. Mining user behavior in a social bookmarking system|a delicious friend recommender system. In International Conference on Data Management Technologies and Applications Data, 2014, 331-338.

[18] Hao Ma. On measuring social friend interest similarities in recommender systems. In International ACM SIGIR Conference on Research and Development in Information Retrieval, 2014, 465-474.

[19] J Fournet and A Barrat. Epidemic risk from friendship network data: equivalence with a nonuniform sampling of contact networks. Scientific Reports, 2016, 6:24593.

[20] V Sekara, A Stopczynski, and S Lehmann. Fundamental structures of dynamic social networks. Proceedings of the National Academy of Sciences of the United States of America, 2016, 113(36):9977.

[21] Mcnett M, Voelker G M. Access and mobility of wireless PDA users. Acm Sigmobile Mobile Computing & Communications Review, 2003, 7(4):55-57.

[22] Yuan P, Liu P. Inter Action Time in Virtual and Physical Spaces: A Comparative Study. International Workshop on Hot Topics in Planet-Scale Mobile Computing and Online Social NETWORKING. ACM, 2015:37-41.

[23] Kaimin Wei, Renyong Duan, Guangzhou Shi, and Ke Xu. Distribution of inter contact time: An analysis-based on social relationships. Journal of Communications and Networks, 2013, 15(5):504-513.

[24] A community resource for archiving wireless at dartmouth; 2007 [cited 2007 august 19 to 26]. database: crawdad [internet]. available from: http://crawdad.org/purpose-user-mobility characterization.html.

[25] Yuan P, Ma H, Fu H. Hotspot-entropy based data forwarding in opportunistic social networks. Pervasive & Mobile Computing, 2015, 16(PA):136-154.