Exploring User Social Behaviors in Mobile Social Applications

Konglin Zhu
Institute of Computer Science
University of Goettingen
Goettingen, Germany
zhu@cs.uni-goettingen.de

Pan Hui
Deutsche Telekom Laboratories/TU-Berlin
Berlin, Germany
pan.hui@telekom.de

Yang Chen
Institute of Computer Science
University of Goettingen
Goettingen, Germany
yang.chen@cs.uni-goettingen.de

Xiaoming Fu
Institute of Computer Science
University of Goettingen
Goettingen, Germany
fu@cs.uni-goettingen.de

Wenzhong Li
Dept. of Computer Science
Nanjing University
Nanjing, China
lwz@dislab.nju.edu.cn

ABSTRACT

Mobile social applications are popular as the proliferating of mobile devices. Understanding user social behaviors is important to improve mobile social applications and enhance its quality of service. However, there is still lack of data for real deployment mobile social application on data analysis of human interaction and social behaviors in mobile social networks.

In this paper, we introduce the experiment methodology of deploying the Goose software in two campuses located in Germany and China respectively. Goose is a mobile social network application allows microblogging, message sending. With the help of volunteers, we collect user interaction data in the duration of 15 days. Based on the collected data, our observation reveals the following aspects of user interactions and their influences. First, user overall activities approximately match user daily life work pattern with a slightly longer time duration and periodically appearance. Second, user encounters in mobile social network follow the heavy tail distribution in small social communities, and user interactions follow the Pareto principle, where about 20% of users make close connections to the other users. Third, communication path between a pair of mobile nodes is mostly within 6 hops, and information diffusion using an epidemic strategy demonstrates that the informed population reaches to 50% in a short term and approaches to 80% in a long term.

Keywords

Mobile Social Networks, User Social Behaviors, Delay Tolerant Networks, Information Diffusion

1. INTRODUCTION

Mobile phones are popular with more than 4 billion phones in use worldwide [1]. Smartphones with programmable capability are growing fraction of these phones. For instance, there are 172 million of smartphones of the 1.2 billion phone sold in 2009 [2]. Meanwhile, social networks have become a popular way for users to interconnect, share contents and express themselves. Popular social networks like Facebook have hundreds of millions of registers and are growing at a rapid pace. As the proliferating of smartphones and social networks, many mobile social applications have been developed to meet social requests of mobile users. For instance, PeopleNet [3] is a mobile social application that seeks information in a mobile social network environment. Prism [4] is developed for remote sensing in a mobile society. The Reality Mining experiment [10] in MIT captured communication, proximity, location and activity information from 100 subjects. However, there is still lack of data for real deployment on mobile social applications on analysis of human interaction and social behaviors in mobile social networks.

Motivated by data collection and human behavior analysis in mobile social networks, we introduce Goose application [11]. Goose is a cellphone application implemented in Nokia Symbian system to support user interactions (i.e. microblogging and messaging) in mobile social networks. With the help of volunteers, we deployed the Goose software in University of Goettingen (UGoe) in Germany and Nanjing University (NJU) in China for data collection. Each university recorded a 15-day log for user activities. Based on the data collected, we investigated user social behaviors in small communities.

Specifically, we make the following contributions in this paper.

- We devise the Goose mobile social application, and propose the experiment methodology of real deployment of the mobile social application in two campuses. We collect user interaction trace during 15 days to analyze human interactions and their influences on mobile social applications.
- We study user overall behavior and find out that user active strength approximately match user daily life work pattern with a sightly longer time duration and periodically appearance.
- We analyze user social interactions. It presents that the number of user encounters in mobile social network follow the heavy tail distribution and the number of user interactions

SNS'11, April 10, 2011, Salzburg, Austria.

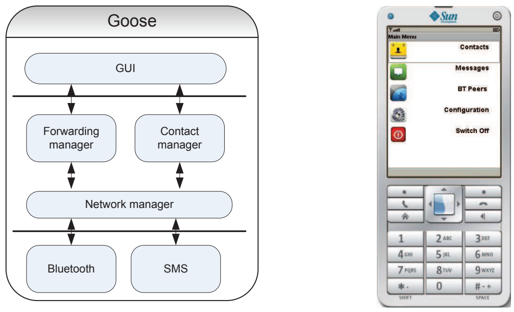


Figure 1: Goose architecture and user interface

follow the Pareto principle, where about 20% of usrs make close connections to the other users.

- We investigate the information diffusion process in mobile social network, which demonstrates that most of communication paths between a pair of mobile nodes are within 6 hops, and information diffusion using an epidemic strategy reaches a half population in a short term and approaches to 80% in a long term.

Although our datasets cannot be more extensive due to limited budget and platforms, we believe this measurement results offer us one step further towards a better understanding of the human mobility pattern and human interaction behaviors in the campus environments.

2. MOBILE SOCIAL NETWORK APPLICATION: GOOSE

Goose is a mobile social network application implemented in Nokia Symbian system to satisfy the low end mobiles social communication among users. It includes various social networking features like microblogging and messaging. It supports one to one communication like sending a text message or voice message to a friend, and one to many communication such as broadcasting news to a group of people by using common wireless network facilities, i.e. GSM and Bluetooth.

2.1 System Design

The Goose architecture contains three major components: contact manager, forwarding manager and network manager, as illustrated in Fig. 1.

Contact manager: it stores and manages user contacts. Every contact entry includes user basic information (e.g. name, phone number and etc.). It records the Bluetooth MAC address for the communication via Bluetooth, and exchange profiles while devices encounter.

Forwarding manager: it is used for message forwarding. It maintains a unique ID for each message and a list of MAC address that already received this message to avoid retransmission. It also takes into account the message priority and available resources to decide whether to transmit information to a dedicated person or not.

Network manager: it coordinates both Bluetooth and GSM network to send and receive data. It records the cellular ID of the current radio tower, detects nearby Bluetooth devices and send friends request to buildup social relatinship. It also take function of switching between different networks based on the connection situation and user preferences.

2.2 Data Transmission

Data transmission in Goose supports unicast, multicast and broadcast mode. Short messages and contact information are sent via unicast mode. Friendship searching information is sent by multicast. Additionally, news and microblogs are sent in broadcast mode.

Broadcast mode only works on Bluetooth network, and the message is sent to the whole network using the epidemic routing strategy [8], which takes the store and forward algorithm for message delivery. Each node stores messages to be sent in its buffer, carries them, and sends the message to the nodes when it encounters, basing on the Bubble Rap algorithm [16].

Unicast messages are delivered via direct forwarding strategy. The text message having a higher priority is sent over GSM network (i.e. SMS) if there is no direct Bluetooth connection between users. Contact information is also exchanged in unicast mode via Bluetooth network when people encounter each other. Furthermore, to update the user profile of an existing contact, a unicast mode profile update request is sent to the targeted user. The targeted user receives the update request and sends the latest user profile to the requester. The multicast message (i.e., friend search) only transmitted to nearby address included in the contact list. This takes inspiration from the fact that someone's details can be easily obtained from a mutual contact.

3. EXPERIMENTS

Goose application is deployed on the campus of UGoe in Germany and the campus of NJU in China to collect data for our analysis of mobile social networks. It targets at investigating human social behaviors in a small community like a university campus.

3.1 Deployment Issues

Software deployment on mobile devices faces the challenges like platform limitations, software capabilities, user incentives and privacy concern in our case.

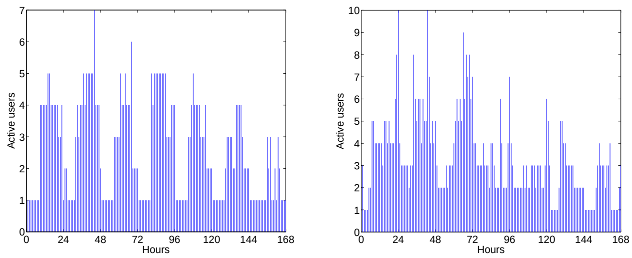
We made the following efforts to attract more people to attend our experiments.

- **Platform extensions:** The Goose application is written by J2ME, which is widely deployed on mobile devices. However, different mobile manufacturers still provide different APIs for their products. It is hardly to provide a platform to support for all different platforms. Instead, we developed different releases for multiple Nokia platforms.

- **Software robustness:** The design of Goose regards to Bluetooth detection and communication. The real test cased cannot perform as robust as the logical design. We well programmed and performed enough tests to guarantee its robustness.

- **User incentives:** One of the most challenges in this experiment is to find volunteers for participating our deployment. We designed a user friendly interface and explain users' concerns (i.e. user privacy) to potential participants. Additionally, we provided small gifts to them to attract people to join the experiment.

- **Privacy concern:** To protect the user privacy, we collected the user interaction data anonymously, which does not contain user basic profiles. We did not log any contents of user interactions. Instead, we only recorded the interaction time, recipients hashing ID, and interaction type. Meanwhile, we hide the basic information of users' contacts in our records to avert the leak of user privacy.



(a) Activity pattern in UGoe (b) Activity pattern in NJU

Figure 2: Activity pattern

3.2 Data Collection

The whole experiment lasts for 15 days in UGoe and in NJU respectively. At the end of the experiment, we collected data from 12 users in UGoe and 15 users in NJU. The software records the logs of user social communications. Specifically, every time when Goose is turned on, the software records the series of {phone model, Bluetooth MAC address, time, Cellular ID}. Furthermore, the log records cellular ID, nearby devices and nearby Goose devices every 2 minutes. To investigate the user interactions, we record the incoming events and outgoing events within this software. For each incoming event, we store a record in the form of {message type, time received, sender, previous relays, message size}. In the experiment, the incoming events include unicast message, broadcast message, status update, contact profile request and acknowledgement, profile update and acknowledgement. The outgoing events only contain unicast message, broadcast message and status update. A record of outgoing message message type, time sent, receiver, message size is logged when a user sending a message.

4. USER SOCIAL BEHAVIOR ANALYSIS

In this section, we discuss the user social behavior in aspect of user overall behavior, user social behavior and information diffusion in mobile social networks.

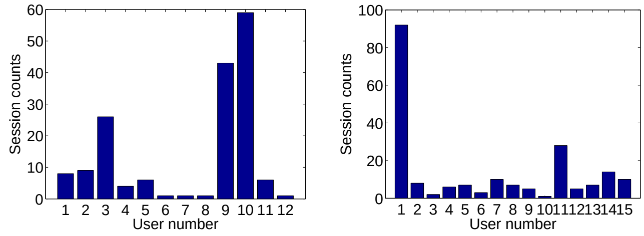
4.1 User Overall Behavior

First, we investigate several user overall behaviors to outline the user activities, mobility and interactions on two campuses.

4.1.1 *User Activity*

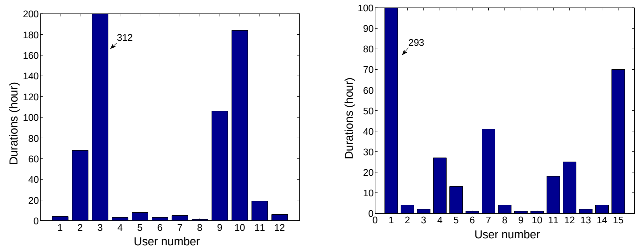
People in our daily life usually follow a certain pattern for each day. For instance, a person stay at office from 9am to 5pm every day. The user activity duration takes a very important role on the applications based on mobility model, such as DTN network [14], information diffusion [15]. These applications use the activity pattern for data dissemination, mobility prediction and etc. Fig. 2a and 2b show one week active users in two campuses. Both figures take 24 hours (i.e., a day) as a unit, and are chronological from Monday to Sunday. Similar as daily life work pattern, users are more active during active time (i.e. from 9am to 12am) than inactive time (i.e. from 1am to 8am). The active length in UGoe is from 9am to 10pm each day. Compared with users in UGoe, the active length in NJU is from 9am to 12am, which is longer than the active length in UGoe. The activity patterns also show that users in weekdays are more active than in weekends. This is attribute to the student activity pattern on campus.

4.1.2 *User sessions*



(a) user sessions in UGoe (b) user sessions in NJU

Figure 3: Session counts



(a) user duration in UGoe (b) user duration in NJU

Figure 4: User duration

The frequency of a user using a typical service is important for service providers to understand and improve the quality of the service. In our experiment, we use sessions to evaluate the frequency of using Goose service. A session is the time difference between the switch on and switch off Goose. Fig. 3a and 3b show the number of sessions each devices initiated during the experiments in UGoe and NJU respectively. Some users are extremely active in the experiments, who initiated 59 sessions in UGoe and 92 sessions in NJU. Meanwhile, there are several users not active in both campuses. It is clear that many of users were not relying upon the Goose application, and they initiated on the average of one session in two days. To understand user activity duration of users, Fig. 4a and 4b depict the duration length of each users in UGoe and NJU respectively. Both campuses shows that only a few of users take a very long time on Goose application. Those most active users spent around 300 hours running Goose application in both UGoe and NJU. Furthermore, more sessions cannot suggest more time duration. Several users run Goose in a long time but only initiated very few sessions.

4.1.3 *User Mobility on Campus*

In order to analyze user mobility on campus, we kept tracking user location during 15 days. The location is estimated by the cellular ID of the radio tower where the user's cellphone locates. Fig. 5 shows the histogram of duration over cells of a typical user. As illustrated in the figure, the user bypasses 29 cells in total during the experiments. Most of the user's location is within the coverage area of 5 cells, which are numbered 1, 2, 3, 10, and 18. Furthermore, in about 80% of the time the user communicates with these 5 cells. This reviews the user's most frequent activity locations where he studies, works and sleeps.

4.1.4 *Messages Statistics*

There are two kind of information collected in our experiment:

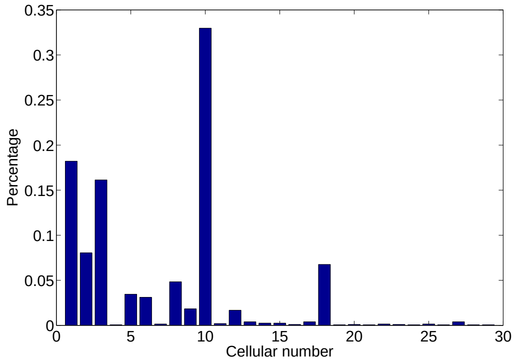


Figure 5: Percentage of a user's time duration

control information and communication information. The control information is used to build up social relationships. And the communication information contains the meta data and content sent/received by users.

The control information consists of 5 kinds of messages.

- **Contact exchange**: the message sent when a user want to add a nearby user into his contact list;
- **Contact exchange acknowledgement**: the message sent to acknowledge the contact information;
- **profile exchange**: the message sent when a user want to update a existing contact's profile;
- **profile exchange acknowledgement**: the message sent to acknowledge the profile exchange message;
- **friend search**: the message sent to nearby friends to seek a unknown contact from a its contact list.

The communication information includes 3 kinds of messages.

- **Unicast message**: the message sent via unicast mode;
- **Broadcast message**: the message sent via broadcast mode;
- **Status update**: the message (i.e. microblogging) sent to friends in broadcast mode;

The percentage of the above messages in UGoe and NJU are shown in Fig. 6a and Fig. 6b respectively. We can see that the percentage of communication information is 59% in UGeo and 87% in NJU. The percentage of control information is lower than the communication information, which indicates that people intend to communicate after they establish social relationship with each other. We can also find that contact exchange messages are more than contact exchange acknowledgements, and profile update messages are more than profile update messages acknowledgements. This attributes to two reasons. Firstly, the exchange or update messages have not been delivered to the counterparts due to disconnection. Secondly, people who do not want to establish or update the relationship with the requester would not acknowledge the request. Fig. 6a shows that unicast messages, broadcast messages and status updates almost have the same percentage, which indicates that users communicate with each other by the three means equally in UGoe. On the other hand, Fig. 6b shows that the broadcast message and status update are more popular than unicast, which indicates people in NJU tends to use microblogging in social communications.

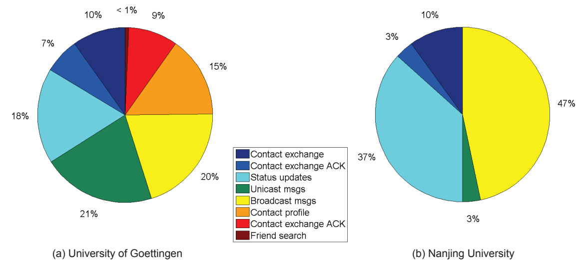


Figure 6: Messages statistics

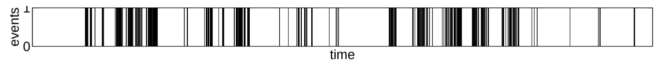
4.2 User Social Behavior

User social behavior has been observed in online social networks [5, 7]. In our experiments, we investigates user social behavior in mobile social network environment, including the heavy tail distribution in user encounters and Pareto principle in user interactions.

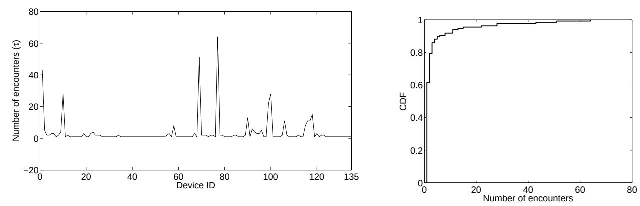
4.2.1 *Heavy Tail Distribution in User Encounters*

Human social activities, ranging from making phone calls to engaging in entertainments, has been reported to appear regularities and follow some patterns [12]. The heavy tail distribution has been observed in complex networks, which is also known as scale-free networks including Internet, WWW, email sending, and paper citations [9]. In a scale-free network, the degree of a node and its distribution is observed to follow the heavy tail distribution.

In our experiment, we study the distribution of user encounters. If two mobile devices are within each other's Bluetooth communication range, it is recorded as an encounter event. Fig. 7a shows the encounter events in UGoe over time scale. It can be seen that the encounter events are dense in some period and sparse in other period. The number of encounters of each device is illustrated in Fig. 7b, which shows that there are high volume of devices encountered only once and quite few devices encountered more than ten times. The cumulative distribution function (CDF) of encounter times is shown in Fig. 7a, which is a heavy tail distribution. Similar distribution is observed in the NJU data set in Fig. 8.



(a) Encounter events



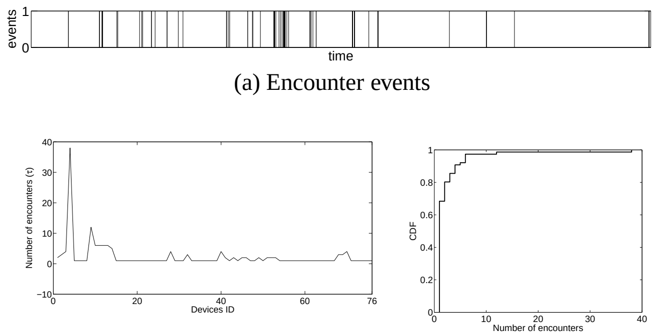
(b) Number of encounters for each device (c) CDF

Figure 7: Encounter behaviors of Goettingen

4.2.2 *Pareto Principle in User Interactions*

User interaction is also an important property in social networks [5]. In our experiment, user interaction is referred to the message received.

The cumulative distribution function (CDF) of user interactions is illustrated in the dashed line in Fig. 9. According to this figure,



(a) Encounter events

(b) Number of encounters for each device

(c) CDF

Figure 8: Encounter behaviors in Nanjing University

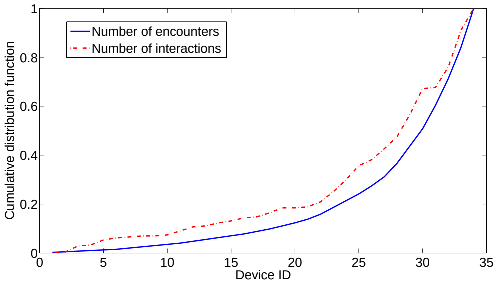


Figure 9: User interactions

most of the users are not socially active: there are about 25 users received messages less than 20%, and only a small fraction of users receive a large part of messages. The CDF of user encounters is also shown in the solid line in Fig. 9. Comparing their curves, we can draw the following conclusions: (1) User interaction behavior roughly satisfies the Pareto principle, which is also known as 80-20 rule, where 80% of the effect comes from 20% of causes (2) User encounter behavior also follows the Pareto principle, namely, 80% of number of encounters comes from 20% of users. (3) There is correlation between user encounters and user inactions, as shown in the figure, the top 20% interactive users are coincidence with the top encountered users.

In summary, there exists a small group of socially active people who make close connections to the others in mobile social networks.

4.3 Information Diffusion in Mobile Social Networks

In our experiments, we employ a delay tolerant epidemic routing strategy to broadcast messages. In this section, we discuss the efficiency of delay tolerant networks (DTNs) routing mechanism and information diffusion process.

4.3.1 DTN Routing Efficiency

The DTN routing efficiency is evaluated by the delivery ratio and the message delay. TABLE I summarizes the messages sent and received during the experiments. It is shown that about 40% of messages are not delivered due to the sparse connection in DTNs. It indicates that DTN routing is only suitable for transmission of messages that are not urgent. It also shows that broadcast message such as status update is delivered in 2 hops in average.

Fig. 10 depicts the number of hops of each message. According to the figure, most message can be delivered within 6 hops, which

	Status	Unicast	Broadcast
Messages sent	27	32	31
Messages received	40	21	32
Unique messages received	20	21	15

Table 1: Total messages transmitted

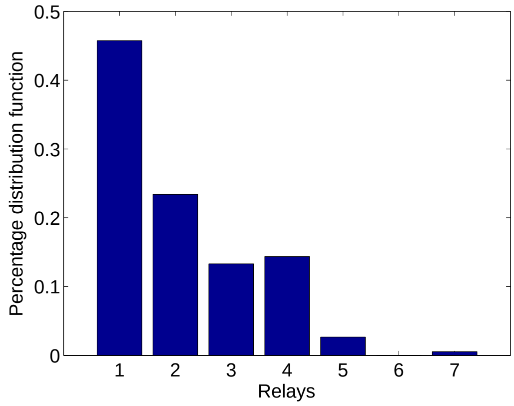


Figure 10: Number of hop of each message

verifies the small world phenomenon studied in [13].

The delay of DTN routing is illustrated in Fig. 11. Delays for broadcast messages and status updates vary from 0 minutes to 10000 minutes. Nearly all of the unicast message delay costs less than 100 minutes. Intuitively, when people send unicast messages to their counterparts via Bluetooth, he usually knows that the proximity of counterpart is in the range of his device. Therefore, most unicast messages in our experiment have short delays. On the other hand, broadcast message needs to be sent to a set of remote destinations, which will introduce longer delay than unicast messages. The long delay of broadcast messaging suggests that application like PeopleNet [3] relying on DTN network cannot be easily deployed in life.

4.3.2 Information Diffusion

Information diffusion has very significant influence in social networks, which has drawn much attention of researchers [7]. Recent study [6] shows that Twitter can be taken as earthquake detectors by taking advantage of fast information diffusion in Twitter. We designed an epidemic scenario when we deploy the software, in which a long life message is broadcasted in our experiments. Fig. 12 depicts the diffusion process of the message. The message is diffused fast before it reaches a half population of nodes. After a half pop-

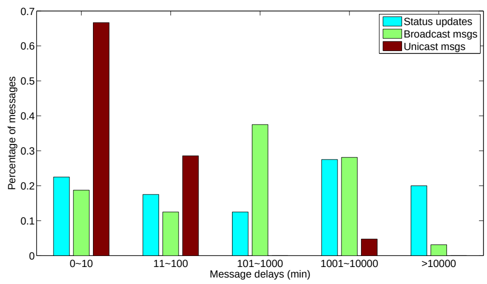


Figure 11: Delay of messages

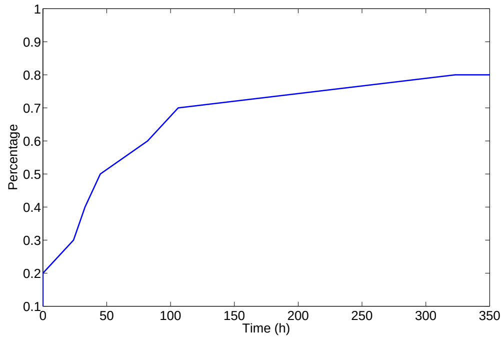


Figure 12: Information diffusion in mobile social networks

ulation received the message, the diffusion speed decreases and finally the message diffusion stopped at 80% of population in the end of the experiment.

In a summary, the information diffuses extremely fast at the beginning of the process, which reaches a half population in a short time. Then the speed of diffusion is slow down and approaches 80% of population in a very long term.

5. RELATED WORKS

5.1 Reality Mining

Reality Mining [10] is one of the largest experiments on mobile platforms. It recorded the user information such as communication, proximity, location, and activity to research on individual and group behaviors in social networks. It addresses the topics such as the evolution of social networks, the predication of people activities, information diffusion, and grouping.

However, Reality Mining did not take much attention on the privacy of users. Even though the anonymity can be finished during data processing, the original data exposed the private data such as user telephone number, user private activities, and etc. Secondly, it is hard to formulate the information diffusion model in Reality Mining dataset.

5.2 IMote Experiment

The IMote experiment [14] consists of several sub experiments to investigate the opportunistic forwarding based on user mobility. Those IMote devices recorded the encounter information of users. By those encountering information, it proposed the design of opportunistic forwarding algorithms. However, the encounter event in IMote experiment is not the real human interaction. Though user encounters provide chances for users to communicate with each other, but it is hard to evaluate the real interaction model only from user encounters.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed the design of Goose mobile social application and its implementation in Symbian system. We deployed the Goose application in two campus and collected data from 27 users in total to analyze the user social behaviors in mobile social networks. The analysis results show that users activity time duration is approximately match the user daily life work pattern, with slightly longer time duration and periodical appearance. User encounters in mobile social network is a scale-free network and conforms the heavy tail distribution in small social groups, and user interactions follows the Pareto principle, where about 20% of users make 80% of interactions in the network. The information diffusion model using an epidemic strategy demonstrates that the informed population reaches to 50% in a short term and approaches to 80% in a long term.

As a part of our project of information diffusion on mobile social networks, we will continue working on the function of Goose application, such as group function and communication with online social networks. We will study the characteristics of information diffusion on mobile social networks in the future.

7. ACKNOWLEDGEMENTS

We would like to thank the volunteers who participated our experiments in both University of Goettingen and Nanjing University. We would also like to thank our anonymous reviewers for their helpful comments.

8. REFERENCES

[1] 4.1 Billion Mobile Subscribers Worldwide Helps Reduce Digital Divide (Slightly), March 2009. http://www.moconews.net/entry/419-4.1-billion-mobilesubscribers-mobile-helping-reduce-digital-divide-sli/.

[2] Worldwide Mobile Phone Sales in 2009, Februrary 2010. http://www.gartner.com/it/page.jsp?id=1306513.

[3] M. Motani, V. Srinivasan, and P. Nuggehalli, PeopleNet: “Engineering a Wireless Virtual Social Network". In Proc. of MobiCom, 2005.

[4] T. Das, P. Mohan, V. N. Padmanabhan, R. Ramjee and A. Sharma, “PRISM: Platform for Remote Sensing using Mobile Smartphones". MobiSys 2010.

[5] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy and B. Y. Zhao, “User Interactions in Social Networks and their Implications”. Eurosys 2009.

[6] U.S. Geological Survey: Twitter Earthquake Detector (TED), September 2010. http://recovery.doi.gov/press/us-geological-survey-twitter-earthquake-detector-ted/.

[7] H. Kwak, C. Lee, H. Park and S. Moon, “What is Twitter, a Social Network or a News Media?”. WWW 2010.

[8] A. Chaintreau, P. Hui, et al. Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms. Infocom 2006.

[9] R. Albert and A.L. Barabasi, “Statistical Mechanics of Complex Networks”. Reviews of Modern Physics, 2002.

[10] MIT Media Lab: Reality Mining, September 2010. http://reality.media.mit.edu/.

[11] N. V. Rodriguez, P. Hui and J. Crowcroft, “Has Anyone Seen My Goose? Social Network Services in Developing Regions”. CSE 2009.

[12] A. L. Barabasi, “The Origin of Bursts and Heavy Tails in Human Dynamics”. Nature 2005.

[13] D. J. Watts and S. H. Strogatz, “Collective Dynamics of Small-world Networks”, Nature 1998.

[14] A. Chaintreau , P. Hui , J. Crowcroft , C. Diot , R. Gass and J. Scott, “Impact of human mobility on opportunistic forwarding algorithms”. IEEE Trans. Mob. Comp, 2007.

[15] D. Gruhl , R. Guha , D. Liben-nowell , A. Tomkins “Information Diffusion Through Blogspace”. WWW 2004.

[16] P. Hui, J. Crowcroft, and E. Yoneki. “Bubble rap: Social-based forwarding in delay tolerant networks”. MobiHoc 2008.