8-2010

Messaging behavior modeling in mobile social networks

Byung-Won ON

Ee Peng LIM
Singapore Management University, eplim@smu.edu.sg

Jing JIANG
Singapore Management University, jingjiang@smu.edu.sg

Freddy Tat Chua CHUA

Viet-An NGUYEN

See next page for additional authors

Citation

Author
Byung-Won ON, Ee Peng LIM, Jing JIANG, Freddy Tat Chua CHUA, Viet-An NGUYEN, and Loo Nin
TEOW

Messaging Behavior Modeling in Mobile Social Networks

Conference Paper · September 2010

6 authors, including:

Ee-Peng Lim
Singapore Management University
447 PUBLICATIONS   **9,044** CITATIONS

SEE PROFILE

Freddy Chongtat Chua
Hewlett Packard Labs
27 PUBLICATIONS   **195** CITATIONS

SEE PROFILE

Viet-An Nguyen
University of Maryland, College Park
14 PUBLICATIONS   **514** CITATIONS

SEE PROFILE

Loo-Nin Teow
DSO National Laboratories
37 PUBLICATIONS   **202** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

ABECOS: Agent Based E-Commerce System View project

Diffusion in Social Networks View project

Messaging Behavior Modeling in Mobile Social Networks

Byung-Won On, Ee-Peng Lim, Jing Jiang,
Freddy Chong Tat Chua, Viet-An Nguyen
School of Information Systems
Singapore Management University
Email: {bwon,eplim,jingjiang,freddy.chua.2009,vanguyen}@smu.edu.sg

Loo-Nin Teow
DSO National Laboratories, Singapore
Email: tloonin@dso.org.sg

Abstract—Mobile social networks are gaining popularity with the pervasive use of mobile phones and other handheld devices. In these networks, users maintain friendship links, exchange short messages and share content with one another. In this paper, we study the user behaviors in mobile messaging and friendship linking using the data collected from a large mobile social network service known as myGamma (m.mygamma.com). We distinguish two types of user behaviors: soliciting active responses for an initiated message and responding to an incoming message. We propose various models for the two behaviors also known as *engagingness* and *responsiveness*. Our experiments show that the two behaviors are quite distinct from each other although they may be correlated. We also show that engaging and responsive users enjoy more friendships. Finally, we show that the engaging and responsive users participate more in messaging about major topics.

I. Introduction

In this paper, we study messaging related user behaviors in myGamma (m.mygamma.net), a well established mobile social networking site that supports both friendship links and messaging services. We distinguish two types of user behaviors: soliciting active responses for an initiated message and responding to an incoming message. The behaviors are also known as user **engagingness** and **responsiveness** respectively. Identifying engaging and responsive users can be useful in a variety of applications including viral marketing, targeted advertisement, network surveillance, online surveys, etc. These users are likely to form the core of a social network and play important roles in spreading messages and getting responses. The presence of such users in the network is also an indication of the vibrancy of network.

Our thesis in this paper is that engagingness and responsiveness behaviors are related to the social status of users in a friendship network as well as their communication patterns with other users. We specifically aim to answer the following interesting research questions: (a) How can we tell if a user is engaging or responsive from his/her messaging activities? (b) How are a user's engagingness and responsiveness behaviors related to his/her status in friendship networks? (c) Are the messaging behaviors related to topics of messages? If so, what are the relationships like?

To verify our thesis and to answer the above questions, models to characterize user engagingness and responsiveness

behaviors are required. Instead of conducting interviews or surveys on users which are more intrusive, costly and time consuming, we define the models using past messages among users. We believe that quantitative models of messaging behaviors should be highly indicative if there are sufficient message data about the users. With the behavior models in place, we proceed to investigate the relationship between messaging behaviors and social status of users measured by number of bi-directed friends. Finally, we seek to uncover the relationships between user engagingness (and responsiveness) and messaging topics.

Modeling user behaviors can be challenging attributed to the wide variety of messages and the connectedness among users in the messaging networks. Messages can be categorized in numerous ways based on its formality, sentiments, and content. Instead of applying natural language text understanding techniques on the message content which is usually computationally costly and inaccurate, we want our messaging behavior models to be defined upon the messaging header data already available as well as the ways (friendship links) users are linked to one another. As one's behaviors can be affected by all his/her neighbors, the messaging behavior models should be able to cope with all the inter-dependency between behaviors.

Mobile messaging in many ways are similar to instant messaging popular among web users. Both support real-time synchronous communications whenever users are online. Mobile messaging however has the additional feature of storing incoming messages whenever users are offline so that the messages can be read when the users become online again. Such a feature enables mobile messaging to behave like email messaging which supports mainly asynchronous communications. As noted in [5], instant messaging users are likely to communicate with few acquainted users as opposed to strangers. Mobile messaging is also different from instant messaging by not restricting the communicating users to be friends on a user's contact list.

The above differences have therefore distinguished our work from the previous works that focus on instant messaging. To the best of our knowledge, engagingness and responsiveness are behaviors yet to be studied in mobile social networks, particularly in large scale. The work presented in this paper is thus early efforts in this direction. Messaging behaviors of

users during online and offline periods can be different yet related. In this paper, we demonstrate that a user's online (and offline) durations can be estimated from the time of messages sent by him/her. From the online durations, we derive the online and offline messaging sessions between users which are in turn used to define the online and offline messaging behaviors.

Our contributions can be summarized as follows:

- We propose several quantitative models for measuring user engagingness and responsiveness in both online and offline messaging sessions. These include the MSGCOUNT, REPLYTIME, SESSIONINIT and SEQUENCE models. We further extend these models to incorporate mutual dependency between engagingness and responsiveness.
- We apply these models on a myGamma dataset containing both messages and friendship links between users. Comparisons between engagingness and responsiveness, and comparisons between different models have been made using this real dataset. We further relate the two behaviors with number of friendships users enjoy.
- We finally show that engaging and responsive users play important roles in messaging topics within an online community. We apply Latent Dirichlet Allocation [2] to uncover latent topics from our message dataset. We discover that major topics in the community are driven by engaging and responsive users.

II. RELATED WORK

There are very few previous efforts on studying user behaviors in email messaging. In [3], user responsiveness behavior is defined in the context of replying emails of the same subject headings. In instant and mobile messaging, message structures are much simpler and subject heading is not longer a viable grouping criteria. This work does not cover the engagingness behavior nor explores different responsiveness behavior models. To the best of our knowledge, there is no other research on modeling messaging behaviors.

As instant messaging is very similar to the myGamma's messaging, we examine related work in the area. Nardi, Whittaker and Bradner found that instant messaging serves largely social purpose instead of formal information exchanges even in the organization setting [5]. Avrahami and Hudson studied the responsiveness of users in instant messaging [1]. The responsiveness here refers to the response time required for a user to respond to an incoming *session initiation attempt* (SIA) message. Strictly speaking, the responsiveness concept here is not a user behavior but some response time label. Unlike [1], we focus mainly on mobile messaging related user behaviors. Due to the peculiar nature of mobile messaging, we have to perform classification of online and offline periods for each user. Instead of treating responsiveness as message response time, we study responsiveness as a quantitative user characteristic. We also introduce engagingness as another user characteristics. Our work is also involved in a much larger dataset.

$SE(u_i)$	Messages sent by user u_i
$RE(u_i)$	Messages received by u_i
$RB(u_i)$	Messages replies sent by u_i
$RT(u_i)$	Messages replying to u_i's earlier messages
OnP_i	Online periods of u_i
$OffP_i$	Offline periods of u_i
$\mathbf{S}_{ij}$	Online sessions between u_i and u_j
$\mathbf{\bar{S}}_{ij}$	Offline sessions between u_i and u_j
$r(m)$	Reply to message m
$Sdr(m)$	Sender of message m
$Rcp(m)$	Recipient of message m
$t(m)$	Sent time of message m
$\mathbf{M}_{i \rightarrow j}$	Messages from u_i to u_j
$\mathbf{M}_{ij}$	Messages between u_i and u_j

III. PRELIMINARIES

Mobile messaging users communicate with one another using a mixture of online and offline messaging sessions. When a user and his/her contact are online, they can exchange messages with each other in real time. On the other hand, a mobile messaging user can also send messages to another user if the latter is offline. In mobile messaging, a mixture of messaging behaviors can exist for the same users. To study these messaging behaviors separately, we first determine these durations automatically based on time gaps between consecutive messages in Section III-A. Once the users' online durations are determined, we proceed to derive the online and offline messaging sessions between every communicating pair of users (see Section III-B).

Table I defines the notations to be used in the rest of paper. A message m' is said to be the *reply* of a m if it is the earliest message that has $Sdr(m') = Rcp(m)$, $Rcp(m') = Sdr(m)$, and $t(m') > t(m)$.

A. Determination of Online and Offline Status

Determining the online and offline communication for mobile messaging users is a non-trivial task. In the absence of a log of user online status over time, we have resort to a statistical approach to automatically decide the online and offline periods of each user as he/she uses the messaging service. Our main proposed idea of segmenting messages into online and offline messages is based on a **Gaussian Mixture Model**. In this model, we envisage that users send messages out at different rates depending on whether they are online or offline. We first define a random variable X for the time gap between two consecutive messages sent by all users. Assume that X is formed by two clusters of time gaps, i.e., online and offline. X can be modeled by a mixture of two Gaussian distributions $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$ where μ_1 and μ_2 represent the mean time gaps of the two distributions respectively, while σ_1 and σ_2 represent the standard deviations respectively. Using EM algorithm, we learn these parameters that generate distributions fitting our dataset. Once the parameters are learnt, the Gaussian distribution with smaller μ_k models the time gaps between sending messages when users are in online periods while another Gaussian distribution models the time gaps when users are in offline periods. We also derive a *time gap threshold* γ to easily classify time gaps into online and offline periods.

B. Online and Offline Sessions

A message session s between two users u_i and u_j is defined by a set of consecutive messages between them. Due to the different online and offline messaging behaviors, we further divide sessions into online and offline sessions.

Given a set of messages $\mathbf{M}_{ij}$ between u_i and u_j, and the online periods of u_i and u_j denoted by $OnP_i = \{[ts_{i1}, te_{i1}], \cdots, [ts_{ik_i}, te_{ik_i}]\}$ and $OnP_j = \{[ts_{j1}, te_{j1}], \cdots, [ts_{jk_j}, te_{jk_j}]\}$ respectively.

The set of overlapping online periods between u_i and u_j, P_{ij}, is defined by:

$$
\begin{aligned}
OlpP_{ij} &= OnP_i \cap OnP_j \\
&= \{[max(ts_i, ts_j), min(te_i, te_j)] | [ts_i, te_i] \in OnP_i, \\
&\quad [ts_j, te_j] \in OnP_j, (ts_i > te_j) \wedge (ts_j > te_i)\}
\end{aligned}
$$

The set of online sessions between u_i and u_j, $\mathbf{S}_{ij}$, is then defined as a collection of message sets induced by the overlapping online periods such that each message set consists of at least some exchange of messages between u_i and u_j.

$$
\begin{aligned}
\mathbf{S}_{ij} = \{&\mathbf{M}_{ij}(p) | p \in OlpP_{ij} \wedge \\
&(\exists m, m' \in \mathbf{M}_{ij}(p), m' = r(m))\}
\end{aligned}
$$

where $\mathbf{M}_{ij}(p) = \{m \in \mathbf{M}_{ij} | t(m) \in p\}$.

The set of online session intervals between u_i and u_j, $OnSsnP_{ij}$, is thus the set of overlapping online periods that cover online sessions, i.e.:

$$
OnSsnP_{ij} = \{p \in OlpP_{ij} | \exists m, m' \in \mathbf{M}_{ij}(p), m' = r(m)\}
$$

From the online session intervals, we derive the remaining periods as:

$$
RemP_{ij} = [min(ts_i^*, ts_j^*), max(te_i^*, ts_j^*)] - OnSsnP_{ij}
$$

where ts_i^* (ts_j^*) and te_i^* (te_j^*) denote the minimum ts_i (ts_j) and maximum te_i (te_j) respectively, in OnP_i (OnP_j).

The set of offline sessions $\bar{\mathbf{S}}_{ij}$ is then defined as a collection of message sets induced by the remaining periods such that each message set consists of at least some exchange of messages between u_i and u_j.

$$
\begin{aligned}
\bar{\mathbf{S}}_{ij} = \{&\mathbf{M}_{ij}(p) | p \in RemP_{ij} \wedge \\
&(\exists m, m' \in \mathbf{M}_{ij}(p), m' = r(m))\}
\end{aligned}
$$

The set of online session intervals between u_i and u_j, $OffSsnP_{ij}$, is thus the set of remaining periods that cover online sessions, i.e.:

$$
OffSsnP_{ij} = \{p \in RemP_{ij} | \exists m, m' \in \mathbf{M}_{ij}(p), m' = r(m)\}
$$

The start and end times of a session s refer to the times of the first and last messages respectively. The user who sends the first message of s is also known as the *initiator* of the session.

Consider the example shown in Figure 1. Users u_i and u_j have two online periods. The messages directed between them are the ones exchanged between u_i and u_j. The messages directed away from them are sent to other users. Although
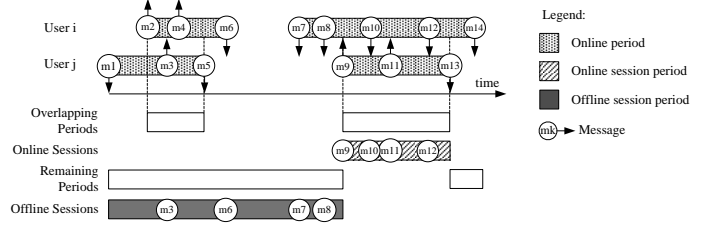

Fig. 1. Online/Offline Periods and Sessions

u_i and u_j are both online in the left overlapping period, it does not constitute an online session due to a lack of message exchange between them. The only online session between u_i and u_j is thus $\{m_9, m_{10}, m_{11}, m_{12}\}$. Among the two remaining periods, only the left one has message exchanges between u_i and u_j. Hence, the offline session found is $\{m_3, m_6, m_7, m_8\}$.

IV. Mobile Social Network Dataset

A. Overview of Dataset

In the myGamma mobile social networking site, members interact and form online communities. Most members are young adults between the age of 20 to 30. The myGamma dataset we obtained consists of 194,809 users and 2.7M messages among them within the one-month period from September 8, 2009 to October 9, 2009. In the dataset, the number of friendship links is 1,795,674. The number of online and offline sessions obtained is 5,491 and 66,806 respectively. Each online (offline) session has about 2 messages (3 messages) on average. It turns out that most users tend to initiate and participate in small number of online and offline sessions. The time gap threshold γ obtained is about 4 hours.

V. User Engagingness and Responsiveness

A. Basic Models

In this section, we will introduce four pairs of basic engagingness and responsiveness behavior models, namely MSGCOUNT, REPLYTIME, SESSIONINIT, and SEQUENCE. They are designed based on message, reply time, session and messaging sequence data respectively. Each model assigns an engagingness (responsiveness) score $\in [0, 1]$ to each user, 0 for non-engaging (non-responsive) user and 1 for fully engaging (fully responsive) user. As users may demonstrate different messaging behaviors during online and offline sessions, every model has both online and offline versions. For example, the online and offline session versions of MSGCOUNT are MSGCOUNT$_{on}$ and MSGCOUNT$_{off}$ respectively.

MSGCOUNT Model: This model is designed based on the principle that an engaging user should have most of his/her messages replied by other users, while a responsive user should have most of his/her received messages replied. The engagingness and responsiveness scores, A^{MSGCOUNT} and

R^{MSGCOUNT}, for online and offline sessions are thus defined by:

$$A_x^{\text{MC}}(u_i) = \frac{|RT_x(u_i)|}{|SE_x(u_i)|} \tag{1}$$

$$R_x^{\text{MC}}(u_i) = \frac{|RB_x(u_i)|}{|RE_x(u_i)|} \tag{2}$$

where session type x can be online or offline denoted by on and off respectively.

REPLYTIME Model: Unlike MSGCOUNT, this model examines the reply times of messages to determine user engagingness and responsiveness. An engaging user should have his/her messages quickly replied by others while a responsive user should have received messages quickly replied. Given a message m' which is a reply of message m, i.e., $m' = r(m)$, the *reply time* of m', is $rt(m') = t(m') - t(m)$. The z-normalized reply time $\hat{rt}(m')$ is defined by $\frac{rt(m') - \overline{rt}}{\sigma_{rt}}$ where $\overline{rt}$ and σ_{rt} are the mean and standard deviation of reply time respectively. Now, we define the engagingness and responsiveness of REPLYTIME model as:

$$A_x^{\text{RT}}(u_i) = \frac{1}{|SE_x(u_i)|} \sum_{\substack{m \in SE_x(u_i) \\ m' = r(m)}} f(\hat{rt}(m')) \tag{3}$$

$$R_x^{\text{RT}}(u_i) = \frac{1}{|RE_x(u_i)|} \sum_{\substack{m \in RE_x(u_i) \\ r(m) = m'}} f(\hat{rt}(m')) \tag{4}$$

where

$$f(x) = \frac{e^{-x}}{1 + e^{-x}} \tag{5}$$

The function $f()$ is designed to convert the normalized reply time to the range [0,1] with 0 and 1 representing extreme slow and extreme fast reply times respectively.

SESSIONINIT Model: In this model, we adopt the principle that an engaging user is more likely to initiate messaging sessions for the messages he/she sends out, while a responsive user is more likely to participate in sessions initiated by messages from others. We first denote the number of online/offline session initiating and participating messages of a user u_i by $SsnInitMsg_x(u_i)$ and $SsnMsg_x(u_i)$ respectively. Let $SE_{on}(u_i)$ be the set of messages sent by u_i during the periods in $OlpP$, and $SE_{off}(u_i)$ be the set of messages sent by u_i during the periods in $RemP$. SESSIONINIT Models for engagingness and responsiveness are then defined as:

$$A_x^{\text{SI}}(u_i) = \frac{|SsnInitMsg_x(u_i)|}{|SsnInitMsg_x(u_i)| + |SE_x(u_i) - SsnMsg_x(u_i)|} \tag{6}$$

$$R_x^{\text{SI}}(u_i) = \frac{\sum_j |SsnInitMsg_x(u_j) \cap \mathbf{M}_{j \to i}|}{\sum_j |SsnInitMsg_x(u_j) \cap \mathbf{M}_{j \to i}| + |\mathbf{M}_{j \to i} - SsnMsg_x(u_j)|} \tag{7}$$

where $SsnInitMsg_x(u_j) \cap \mathbf{M}_{j \to i}$ represents the set of messages from u_j to u_i that successfully initiate online (or offline) sessions with u_i, and $\mathbf{M}_{j \to i} - SsnMsg_x(u_j)$ represents the set of messages from u_j to u_i that fails to initiate online (or offline) sessions with u_i.

SEQUENCE Model. Message sequence refers to the sequence of messages sent and received by a user ordered by time. To derive engagingness and responsiveness from message sequences, we consider the principle that an engaging user is expected to have his or her sent messages replied soon after they are received by the message recipient, and a responsive user replies soon after they receive messages. As the time taken to reply an message may vary, we consider the number of messages received later than a message m but are replied before m by a user as a proxy of how soon m is replied.

The above principle is thus used to develop the SEQUENCE Model. Let $seq_{x,i}$ denote the online ($x = on$) or offline ($x = off$) session message sequence of user u_i. When a message received by u_i is replied before other message(s) received earlier, the reply of the former is known as an *out-of-order reply*. Formally, for a message m received by u_i, we define the *number of messages received* and *number of out-of-order replies* between m and its reply m' in $seq_{x,i}$, denoted by $n_{x,r}(u_i, m)$ and $n_{x,\overline{o}}(u_i, m)$ respectively, as

$$
\begin{aligned}
&n_{x,r}(u_i, m) \\
&= \begin{cases} \text{\# messages received between} & \text{if } \exists m' \in RT_x(u_i), \\ \text{m and m' in } seq_{x,i}, & \quad r(m) = m' \\ -1, & \text{otherwise} \end{cases}
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
&n_{\overline{o}}(u_i, m) \\
&= \begin{cases} \text{\# messages received} & \text{if } \exists m' \in RT_x(u_i), \\ \text{between m and m' in } seq_{x,i} & \quad r(m) = m' \\ \text{and have been replied,} & \\ -1, & \text{otherwise} \end{cases}
\end{aligned} \tag{9}
$$

The -1 value is assigned to $n_{x,r}$ and $n_{x,\overline{o}}$ when m is not replied at all. The user engagingness and responsiveness of the SEQUENCE$_x$ model are thus defined as:

$$A_x^{\text{SQ}}(u_i) = \frac{\sum_{m \in SE_x(u_i), u_j = Rcp(m)} \left(1 - \frac{n_{x,\overline{o}}(u_j, m)}{n_{x,r}(u_j, m)}\right)}{|SE_x(u_i)|} \tag{10}$$

$$R_x^{\text{SQ}}(u_i) = \frac{\sum_{m \in RE_x(u_i)} \left(1 - \frac{n_{x,\overline{o}}(u_i, m)}{n_{x,r}(u_i, m)}\right)}{|RE_x(u_i)|} \tag{11}$$

B. Mutual Dependency Based Models

In the above basic models, user engagingness and responsiveness are computed independently. They share the same underlying assumption that messaging behaviors of a user is independent of other users. This assumption does not always hold in practice as user behaviors are likely to be affected by other users he or she communicates with. Hence, we have designed the mutual dependency based engagingness and responsiveness models.

Suppose $A^M(u_i)$ and $R^M(u_i)$ are engagingness and responsiveness of user u_i computed using model M. The mutual dependency between A^M and R^M can be expressed as:

	A^{RT}	A^{SI}	A^{SQ}	$A^{\text{MC*}}$	$A^{\text{RT*}}$	$A^{\text{SI*}}$	$A^{\text{SQ*}}$
A^{MC}	0.86	0.98	0.99	0.86	0.86	0.73	0.86
A^{RT}		0.85	0.86	0.99	0.99	0.79	0.99
A^{SI}			0.98	0.85	0.85	0.75	0.85
A^{SQ}				0.86	0.86	0.74	0.86
$A^{\text{MC*}}$					0.99	0.79	0.99
$A^{\text{RT*}}$						0.79	0.99
$A^{\text{SI*}}$							0.79

	R^{RT}	R^{SI}	R^{SQ}	$R^{\text{MC*}}$	$R^{\text{RT*}}$	$R^{\text{SI*}}$	$R^{\text{SQ*}}$
R^{MC}	0.85	0.98	0.99	0.86	0.85	0.97	0.86
R^{RT}		0.81	0.86	0.99	0.99	0.88	0.99
R^{SI}			0.98	0.81	0.81	0.99	0.81
R^{SQ}				0.86	0.86	0.97	0.86
$R^{\text{MC*}}$					0.99	0.88	0.99
$R^{\text{RT*}}$						0.88	0.99
$R^{\text{SI*}}$							0.88

- A user is considered more engaging if he/she can get less responsive users to respond. Formally, we write:

$$A^{M*}(u_i) = \frac{\sum_{u_j} v^M_{u_i,u_j} \cdot (1 - R^M(u_j))}{|SE_x(u_i)|} \qquad (12)$$

- A user is considered more responsive if he/she responds to less engaging users.

$$R^{M*}(u_i) = \frac{\sum_{u_j} w^M_{u_i,u_j} \cdot (1 - A^M(u_j))}{|RE_x(u_i)|} \qquad (13)$$

where $v^M_{u_i,u_j}$ and $w^M_{u_i,u_j}$ denote the quantity values between u_i and u_j computed based on the principle of M (i.e., # of replies between u_i and u_j in $A^{\text{MC}}_x(u_i)$).

VI. EXPERIMENT RESULTS - COMPARISON OF MESSAGING BEHAVIORS

For comparison between user behavior models, we compare by examining Spearman's rank correlation coefficient. The Spearman's rho of two ranked list l_1 and l_2, $\rho(l_1, l_2)$ is defined by:

$$\rho(l_1, l_2) = 1 - \frac{6 \sum d^2_{u_i}}{n(n^2 - 1)} \qquad (14)$$

where l_1 and l_2 have n users' ranks and the difference $d_{u_i} = l_1(u_i) - l_2(u_i)$ between the ranks of user u_i on l_1 and l_2. ρ value falls between -1 and 1 representing negative correlation and positive correlation respectively. In addition, $\rho = 0$ stands for no linear correlation.

Comparison between user engagingness (responsiveness) models. Table II (Table III) shows the *Spearman's rho* between the ranked lists produced by different engagingness (responsiveness) models for online sessions. The table shows that most

Model	Spearman's rho	Model	Spearman's rho
MC	0.83	$MC*$	0.75
RT	0.75	$RT*$	0.75
SI	0.78	$SI*$	0.72
SQ	0.83	$SQ*$	0.75

engagingness (responsiveness) models are very similar to one another except A^{SI} and $A^{\text{SI*}}$ which are slightly more different. This is because of the principle of the SESSIONINIT Model which is distinct from the other models. In the SESSIONINIT Model, the engagingness of a user will be high when the user tends to initiate a number of sessions. However, it turns out that most users usually initiate a small number of sessions in the myGamma dataset. Though not shown here, we also observe the same for engagingness (responsiveness) in offline sessions.

Comparison between engagingness and responsiveness. Next, we examine the difference between engagingness and responsiveness for different models for online sessions. As shown in Table IV, the Spearman's rho values between the two behaviors of the same model are mostly more different than differences observed between two models for the same behavior (say, engagingness). The only exception is SESSIONINIT model. This can be relatively sparser data for measuring the model. Interestingly, for offline sessions, we observe that the distinction between engagingness and responsiveness is less obvious. This could be due to offline nature (i.e., long time lag) of responding messages between users.

Engagingness/responsiveness and friendship links Figure 2 depicts the boxplots of number of bi-directed friendship links of users divided into five different engagingness/responsiveness intervals of size 0.2. Here, we derive the overall engagingness (responsiveness) of each user by averaging the engagingness (responsiveness) of different models (including online and offline versions). We observe that users with higher engagingness have more friendship links. This is less obvious for responsiveness. This suggests that engaging users are more capable of attracting and establishing friendships.

VII. EXPERIMENT RESULTS - TOPIC SPECIFIC MESSAGING BEHAVIOR ANALYSIS

A. Motivation

Users demonstrate different messaging behaviors in different topics of discussion. For interesting topics, one expect users to be more engaging and responsive, while uninteresting topics will only turn users away from participation. In this section, we analyze user engagingness and responsiveness for different message topics in our dataset. The purpose here is to identify interesting topics within the online community.

To conduct this study, we first identify the major message topics from the aggregated message content for a set of users using Latent Dirichlet Allocation (LDA) [2]. We then analyze the distribution of engagingness and responsiveness of users within each message topic.

B. Message Topic Distillation

For our analysis purpose, we only select users indicating English as their preferred language and there are only 27,920 such users. Despite this pruning effort, there are still some users writing non-English messages as shown in our results. Due to the limited content in each message, we aggregate
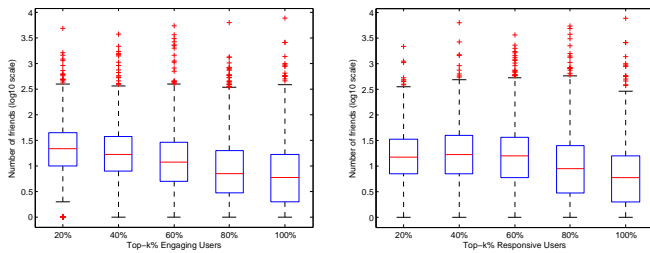
Fig. 2. Engagingness/responsiveness and friendship links.

TABLE V
MAJOR TOPICS.

Topics	Top 10 terms
T14	love, chat, hello, want, dear, baby, friend, dont, hope, miss
T15	dear, chat, sana, sawa, doin, kwani, swty, pliz, thea, sasa,
T17	view, blkapp, mode, click, gift, return, gifts, love, private, thank

(a) Major Topics (b) Non-Major Topics
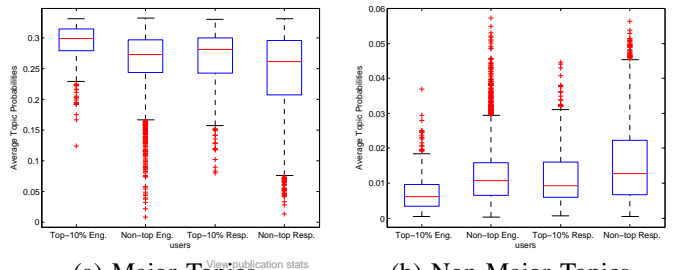
Fig. 3. Average Topic Probability Distribution.

the messages by their senders and recipients. Messages sent by a user capture the topics in which he/she is interested to communicate with others. On the other hand, messages received by a user represent the topics about which others wish to communicate with him/her. We call the two aggregated message content the out-document and in-document of the user. We also remove stop words from these content using a combined dictionary of 400+ stop words from [4]. Given a set of documents and k topics, LDA essentially finds the k latent topics in the documents such that each document is assigned a topic distribution, and each word occurrence in the document is assigned a topic. Since topics are not given beforehand, we performed LDA on the merged set of out-documents and in-documents with $k = 20$ common topics. The empirical choice of $k = 20$ appears to work well as we could find the popular topics exist in the data.

The topic distillation results are shown in Table V. A uniform topic distribution assumption for users would have 0.1 assigned for each topic. Among the 20 topics, most have only a few hundreds of users (e.g., topic 1 has 141 users), while topics 14, 15, and 17 have 27,741, 17,088, and 4,780 users respectively. We call these users the main users. We empirically select topics 14, 15 and 17 as the major topics as they have much more main users. The remaining topics are thus the non-major topics.

To conserve space, we only show the top 10 terms found in the three major topics. Topic 14, the largest topic in term of main user count, consists of mainly greeting terms. This is not a surprise as users tend to greet one another in such a social network. Topic 15 appears to be dominated by abbreviated (e.g., "doin"="doing", "swty"="sweety") and non-English terms (e.g., "sana", "sewa", "kwani"). Topic 17 is likely to be related to use of software and exchange of gifts.

C. Messaging Behaviors in Message Topics

We would now like to examine the distinction between engaging (or responsive) users and other users in both major and non-major topics.

Figure 3a shows the boxplots of top 10% engaging (responsive) users' average major topic probabilities and those of non-top engaging (responsive) users. The average major topic probability of a user is derived by averaging the topic probabilities of his/her out-documents (in-documents) for the major topics (i.e., Topics 14, 15 and 17). Similarly, we derive the average non-major topic probability of each user in Figure 3b. Figure 3a shows that the top 10% engaging users contribute more to the major topics than the other users. On the other hand, the former contribute less on average to the non-major topics than the other users as shown in Figure 3b. From the figures, we also observe the major topics enjoy more user contribution than non-major topics in general. We also examine the average topic probability of top 10% responsive users and non-top 10% responsive users for major topics and non-major topics in Figure 3 showing similar results to engaging users. On the whole, the results match our intuition that engaging and responsive users are the ones driving important topics in the online community. That is, the former tends to generate messages of major topics while the latter tends to receive messages of major topics.

VIII. CONCLUSION

In this paper, we study user engagingness and responsiveness as two messaging behaviors in a mobile social network community. Our experiments on the real dataset show that engagingness and responsiveness are largely distinct during the online sessions but less distinct during the offline ones. We also show that engaging and responsive users enjoy more friendship links and are also the ones dominating major topics found in the messages.

REFERENCES

[1] D. Avrahami and S. E. Hudson. Responsiveness in instant messaging: predictive models supporting inter-personal communication. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 731–740, 2006.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[3] P. Deepak, D. Garg, and V. Varshney. Analysis of Enron Email Threads and Quantification of Employee Responsiveness. In *Workshop on Text Mining and Link Analysis (TextLink 2007)*, 2007.

[4] S. Howard, H. Tang, M. Berry, and D. Martin. GTP: General Text Parser. In *http://www.cs.utk.edu/~lsi/*, 2009.

[5] B. A. Nardi, S. Whittaker, and E. Bradner. Interaction and outeraction: instant messaging in action. In *ACM conference on Computer supported cooperative work (CSCW)*, pages 79–88, 2000.