Predicting Social Networks and Psychological Outcomes Through Mobile Phone Sensing

W. Quin Yow*, Xiaqian Li*, Wan-Yu Hung*, Megan Goldring†, Long Cheng*, Yu Gu*
*Singapore University of Technology and Design, Singapore
†University of Southern California, USA

Abstract—Proliferation of mobile phones over the last decade has led to innovative methods for studying social behavior and friendship patterns. Our present study combined advanced mobile phone technologies, such as Bluetooth scanning and automated pushing surveys, with self-reported questionnaires to examine the social structure and psychological well-being of college students. We distributed smartphones to 35 first-year undergraduate students to use daily over one academic term (three months) and asked them to complete questionnaires on language background, loneliness, adaptation to college life, feelings of community cohesion, and friendship ties. Results suggested that behavioral networks in physical co-location, obtained using our mobile phone sensing techniques, reliably inferred real reported friendships. In addition, mobile phone usage, such as calling and SMS, were significantly correlated with psychological well-being in terms of feelings of loneliness, sense of community cohesion and adaptation to college life. Finally, results revealed that activities in the mobile phone social network, in particular SMS, may influence students’ language use, such that students tended to adapt their language behavior to match that of their SMS partners.

Index Terms—Mobile phone sensing, social factors, social network analysis, wireless ad hoc network.

I. INTRODUCTION

Social Network Analysis (SNA) has been widely used to study complex human relationships in modern sociology and social psychology. Traditional social network studies have relied largely on self-reported data with a limited number of observation points [1], [2]. Mobile phone, however, with its rapid proliferation amongst billions of people, is an ideal computing platform for experimental social psychological research, especially in the study of dynamic social networks. It offers a cost-effective and unobtrusive means to collect information about human behavior [3]. Recent years have seen an emerging trend to use mobile phones as tools to continuously collect directly observable behavioral data (e.g. call logs and physical proximity) [4], [5]. For example, in a study of 94 students and faculty from the same university, Eagle and colleagues found that call data and proximity data from mobile phones could predict as much as 95% of friendships [6]. Hence, mobile sensing is capable of cross-validating the data obtained through traditional methods such as self-reported questionnaires.

Social networks have the potential to influence all aspects of our lives, both psychologically (e.g. job satisfaction [6], happiness and loneliness [7], [8]) and behaviorally (e.g. service adoption [9]). Social network ties are more likely to exist among individuals with similar demographic backgrounds and/or behavior patterns than among dissimilar individuals (a tendency also known as network autocorrelation [10]). Two theories have been proposed to explain this autocorrelation process. The homophily principle states that people tend to connect to similar others because doing so is easier and/or more rewarding [11]. In contrast, the assimilation principle states that dissimilar individuals adjust their own behavior patterns to meet those of their social network partners [12]. However, despite recent proliferation of mobile phone use, it remains unclear what drives the formation of social ties in mobile phone social network, and how such ties influence our psychological well-being and behavior. In this paper, we aim to address this issue by applying advanced mobile phone sensing technologies and automated pushing surveys to examine how interpersonal interactions via phone calls, text messages (i.e. SMS) and physical co-location among first-year undergraduate students may affect their self-reported friendships, psychological well-being and language use over the first term of their university life.

Our study seeks to explore whether: 1) friendship networks can be inferred from mobile phone data (e.g. physical co-location), 2) mobile phone usage (call and SMS) influences students’ psychological well-being, such as feelings of loneliness, sense of community and adjustment to college life, and 3) mobile phone social networks co-evolve with students’ language use in communication, in particular, their code-switching behavior (switching between two or more languages in the context of a single conversation), as language mixing is a common phenomenon amongst the undergraduate students.

The remainder of this paper is organized as follows: Section II describes the methodology. Section III provides the results of the study followed by discussion in Section IV. Finally, Section V concludes the paper.

II. METHODOLOGY

Thirty-five first-year students (19 males, 16 females) at a local university in Singapore participated in this study. Of the 35 participants, 22 were local Singaporeans while the remaining
13 were international students from China (5), Malaysia (4), India (2) and Vietnam (2). The participants were from two different cohort classes studying on the same campus and staying in the same dormitory. We asked the participants to use Android smartphones preinstalled with our “Social Statistics” software that could record and send phone usage data and co-location information to our server located in the university. We also collected self-reported data from questionnaires on language background, friendship tie-strength, loneliness, classroom community, and adaptation to college life.

A. Data Collection Through Mobile Phone Sensing

Our system employed standard client-server architecture, as shown in Fig. 1. The software application of client ran on any android mobile phone, and would mandatorily start at bootup. To avoid data loss for a long time, a daemon monitor reported whether the application was running or not on the phone. For energy saving purpose, the application was turned off from 12:00 am to 7:00 am. Moreover, phones only used the free campus WiFi network to upload collected data. Data were cached locally and uploaded once campus WiFi connection was detected. Since all participants were first-year students and were typically on campus during weekdays, the local storage space (16GB) was enough to cache data for several days. From our experimental results, each phone had the campus WiFi connection for 5.17 hours every day on average.

By default, data were sampled once every 5 minutes. We also provided user interface to adjust the data sampling rate. A sensing report consisted of three parts, as shown in Fig. 2, namely the logical connection, physical connection, and contextual information. More specifically, the logical connection included the phone call and SMS, which could be obtained from the phone usage log. To protect user privacy in the mobile phone sensing, we only recorded the statistics on phone usage, such as call frequency and duration. We used the WiFi fingerprinting method for indoor localization, which achieved an accuracy of 5-7 meters in our study [13]. We used the periodic Bluetooth scanning to discover co-located participants. For the purpose of building a comprehensive mobile sensing dataset, we also collected most of the available sensory data on the smartphone, such as magnetic, audio, accelerometer and gyroscope sensor data.

B. Psychological Measurements

We administered the same set of questionnaires at both the start (i.e. T1, May 2013) and the end (i.e. T2, Aug 2013) of the first academic term with an interval of approximately 13 weeks:

- Language Background and Code-switching Questionnaire was used to collect information about participants’ language experiences and their code-switching behavior. Participants indicated their proficiency and usage of the language(s) they knew and the frequency of code-switching in the past week under different contexts (e.g. face-to-face communication, text messaging, academic setting, casual setting, etc.). In addition, participants received the same set of code-switching questions automatically from their phones every two weeks and were asked to submit their answers via their phones. Thus, code-switching data were collected at seven different time points in total, with an interval of about 10-12 days.

- Friendship Tie-strength Questionnaire [14] was used to assess self-reported friendship networks among the participants. Participants indicated their tie-strength with each other on a 9-point scale, e.g. “How strong is your relationship with this person”, where 1 indicated they barely knew each other and 9 indicated they were close to each other.

- UCLA Loneliness Scale (Version 3) was used to measure loneliness in college students [15]. It consisted of 20 items and the total score ranged from a minimum of 20 to a maximum of 80, where a higher score indicated a greater sense of loneliness.

- Classroom Community Scale consisted of 20 items that measured individual sense of community in a learning environment [16]. The total score ranged from 0 to 40 with a higher score indicating a greater sense of community.

- Student Adaptation to College Questionnaire (SACQ) was used to measure college adjustment [17]. It consisted of four subscales (Academic Adjustment, Social Adjustment, Personal-Emotional Adjustment and Goal Commitment/ Institutional Attachment) with a total of 67 items. The total score ranged from 67 to 603, and higher scores indicated better adjustment.

III. RESULTS

A. Technical Data

Descriptive statistics on phone usage data (call and SMS) are shown in Table I and Table II. On average, each participant made 441 times of phone calls and spent 21986s in total (i.e.
366.43 minutes) communicating on their mobile phones, and sent/received 767 text messages during the whole period of the study.

Co-location data were generated by periodic Bluetooth scans at 5-minute intervals. For each co-location event, we logged the two co-located smartphones and the time and date of the event. Because all phones were scanning every five minutes, if two participants were together for 100 minutes, there would be 20 recorded co-location events for each participant, which would be a total of 40 co-location events for the dyad. We therefore approximated each co-location event to be representative of a 2.5-minute time interval for one dyad. We detected 270,000 co-location events of all 35 participants (i.e. 454 co-location events for each dyad), where each dyad was together for about 1,135 minutes. When we examined whether the co-location events happened during weekdays or weekends, we found that 98% of the co-location events happened during weekdays. Weekdays were defined as Monday 12 am to Friday 5 pm in this study. This is likely due to the fact that most local students went home and stayed with their families during weekends.

In order to examine the relationship between self-reported friendship data and observed behavioral data (information obtained from mobile phone sensing, e.g. physical co-location), we generated three inferred friendship networks based on co-location data over three different periods: first four weeks, last four weeks, and overall co-location (see Fig. 4). For these three inferred friendship networks, larger nodes depicted participants who shared more co-location events with all other participants and a link between two nodes represented the number of co-location events detected between the dyad, with a link in a darker color indicating more time spent co-located with each other. Nodes were physically separated by modularity class and node colors highlighted the different communities in a network. A link from a node A to a node B represented the tie-strength reported by A to indicate his/her friendship with B and a link in a darker color indicated a higher tie-strength score. In addition, participants were divided into communities (sub-networks) based on modularity class [19] and each community was represented by a different color. Table I shows the inferred friendship networks only using co-location data over the first and last four weeks (modularity$_{T1}$=0.156; modularity$_{T2}$=0.217). Each network consisted of two larger communities and one-to-three isolated individuals, who had no co-location events with other participants at all. By using co-location data from the entire term, Fig. 4(c) shows two communities for the entire inferred friendship network (modularity$_{Overall}$=0.211) and the communities exactly corresponded to the cohort classes to which the participants belonged. Comparing graphs in Fig. 3 and Fig. 4, we found that the inferred friendship networks largely replicated the self-reported friendship networks in terms of the community structure within the network. In particular, participants from the same cohort class spent more time together. This is likely due to the university’s unique cohort structure, where first-year students from the same cohort class had the same schedule for lectures and shared one cohort classroom.

B. Inferring Friendship

All participants were asked to indicate their relationship with each of the other participants at T1 and T2. Fig. 3 shows these two reported friendship networks produced by Gephi [18], a visualization tool. In each graph, participants were represented by circles (nodes) and the numbers in the circles depicted which cohort classes the participants were from. Size of the circles corresponded to how much the person was rated as a close friend by every other participant (tie-strength total scores). A link from a node A to a node B represented the tie-strength reported by A to indicate his/her friendship with B and a link with a darker color indicated a higher tie-strength score. In addition, participants were divided into communities (sub-networks) based on modularity class [19] and each community was represented by a different color. Fig. 3(a) shows two communities for the entire self-reported friendship network at T1 (modularity$_{T1}$=0.123). The red community had all participants from cohort class 2, while the cyan community consisted of all participants from cohort class 1. Fig. 3(b) displays two communities for the self-reported friendship network at T2 (modularity$_{T2}$=0.156). Similar to T1, the communities at T2 corresponded roughly to the cohort classes of the participants. Participants were more likely to report others as close friends if these others were also from the same cohort class as themselves than if they were not.

The Quadratic Assignment Procedure (QAP) technique was used to analyze the social network data [20] and it was found that the two reported friendship networks (T1 and T2) were significantly correlated ($r=0.65, p<0.001$; a $p$-value less than 0.05 indicates significance at a statistical level). Self-reported friendship network at T1 was significantly related to self-reported friendship network at T2, implying that these friendship networks were relatively stable throughout the term. We also found higher tie-strength mean scores (an indication of how close two individuals were) in T2 than T1 (mean$_{T1}$=3.06, $SD_{T1}$=2.41; mean$_{T2}$=4.09, $SD_{T2}$=2.66; $t(1190)=16.78, p<0.001, d=1.03$). This suggested that participants within the networks tended to become closer to each other over time.

![Table I](image)

<table>
<thead>
<tr>
<th>Phone Call Usage</th>
<th>Weekdays</th>
<th>Weekends &amp; Weekends</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>162</td>
<td>221</td>
</tr>
<tr>
<td>Duration (sec)</td>
<td>6812</td>
<td>9383</td>
</tr>
</tbody>
</table>

![Table II](image)

<table>
<thead>
<tr>
<th>SMS Usage</th>
<th>Weekdays</th>
<th>Weekends &amp; Weekends</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>311</td>
<td>419</td>
</tr>
<tr>
<td>Length</td>
<td>17657</td>
<td>23971</td>
</tr>
</tbody>
</table>

1Modularity measured how well a network decomposed into modular communities. A modularity score ranged from -0.5 to 1 and a higher score indicated a more sophisticated internal structure, usually with more communities.
(a) Self-reported network at T1
(b) Self-reported network at T2

Fig. 3. Self-reported friendship networks. Nodes (circles) represent participants in the study and links represent tie-strength between two individuals. The number inside each node depicts which cohort class the participant was from (1=cohort class 1; 2=cohort class 2). Node colors indicate different communities in the network (red=modularity class 1; cyan=modularity class 2).

TABLE III

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Self-report T1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Self-report T2</td>
<td>0.65†</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Co-location T1</td>
<td>0.11‡</td>
<td>0.21†</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4 Co-location T2</td>
<td>0.14†</td>
<td>0.16‡</td>
<td>0.56†</td>
<td>-</td>
</tr>
<tr>
<td>5 Co-location Overall</td>
<td>0.16†</td>
<td>0.23‡</td>
<td>0.82†</td>
<td>0.86‡</td>
</tr>
</tbody>
</table>

† Correlation was significant at the 0.001 level.
‡ Correlation was significant at the 0.01 level.
§ Correlation was significant at the 0.05 level.

Importantly, QAP revealed significant positive correlations between reported friendship (T2) and inferred friendship based on co-location data over the first four weeks, the last four weeks and the whole term (r=0.21, r=0.16, r=0.23, respectively, ps<0.01). This suggested that we were able to infer friendships based on observable measures like co-location using mobile phone sensing technologies.

C. Loneliness and College Adjustment

The UCLA Loneliness Scale (LS), Classroom Community Scale (CCS) and Student Adaptation to College Questionnaire (SACQ) all had high internal reliabilities. Cronbach’s αs for the current study were as follows: αLS = 0.85 (T1), 0.90 (T2); αCCS = 0.90 (T1 and T2); αSACQ = 0.86 (T1), 0.96 (T2). Paired sample t-test indicated no difference between T1 and T2 on total scores for all the three questionnaires (ps>0.05), implying that participants’ responses to the above three scales did not change over time. Hence, only data from T2 were used to further investigate the relationship between mobile phone usage and individual psychological outcomes.

First, we found that responses to these three questionnaires were significantly correlated with each other. Sense of classroom community was positively correlated with college adaptation (r=0.83, p<0.001) and loneliness was negatively related with both classroom community and college adaptation (rs<−0.47, ps<0.01). In summary, participants who had a greater sense of classroom community and better adaptation to their college life were less likely to feel lonely.

Second, when responses to the three questionnaires were compared with phone usage (both phone calls and SMS), we found that overall frequency of all received phone calls was positively correlated with Classroom Community score (r=0.35, p<0.05), and Social Adjustment, one of the subscales in SACQ (r=0.34, p<0.05). Thus, participants who received a greater amount of phone calls also tended to have a greater sense of class community and better social adjustment in college. This suggested that mobile phone usage may play an important role in maintaining social cohesion and psychological well-being. Similar results were found for received calls during weekdays but not for received calls during weekends. Participants who received more phone calls during weekdays were less likely to feel isolated in the college environment and more likely to be coping well with the social demands inherent in the college experience.

D. Co-evolution of Language Use and Social Networks

We investigated the co-evolution of mobile phone social networks, physical co-location, and code-switching use using RSienna, a stochastic actor-driven modelling technique for repeated measures of social networks [21]–[23]. Here we present three stochastic models for phone calls, SMS and co-location networks respectively. Each of the three models comprised two parts. The first part modelled network dynamics by considering...
effects from the network structure and participants’ attributes and behavior. These model parameters were described in Table IV. Gender, cohort class, nationality and code-switching use were included as participant attributes/behavior for modelling homophily effects and how these attributes/behavior might influence the participants’ popularity and activity in the networks. The second part modelled code-switching dynamics by considering effects from the network dynamics, such as assimilation effects, as well as effects from gender, cohort class and nationality.

In the model estimation for call network dynamics (see Table V), a significant negative outdegree was found. This indicated that participants were selective in whom they called (as opposed to randomly calling others) and their choice of call partners could be predicted from other parameters in the model (p<0.001). Specifically, the call choices were significantly reciprocal and tended toward triadic closure (p<0.001). There were significant homophily effects of gender, cohort class, and nationality (all p<0.05), implying that participants tended to call those who were of the same nationality and gender and from the same cohort classes as themselves. A marginally significant indegree-related popularity (p=0.077) indicated that popularity in call network (i.e. receiving more calls) had a self-reinforcing effect over time. The participants were more likely to call popular partners than non-popular partners, such that the popular partners tended to become even more popular over time (a phenomenon also known as the Matthew Effect in sociology that describes a tendency for the rich to become richer [23, 24]). In addition, girls were marginally more popular than boys in the call network (p=0.053). No other significant effects were found.

In the model estimation for SMS network (see Table VI), a significant negative outdegree effect was also found, indicating that participants were selective in whom they sent text messages to (p<0.001). Their choices of SMS partners were significantly reciprocal as in the call network (p<0.001). However, unlike the call network, the SMS network did not show a significant tendency toward triadic closure. There was a significant homophily effect of cohort class (p<0.01), implying that participants tended to SMS those who were from the same cohort class, but only a marginal effect of nationality (p=0.058). There was no indegree-related popularity effect, suggesting that participants were as likely to SMS non-popular partners as popular partners. Most interestingly, code-switching use was found to have a significant influence on SMS activities,

![Inferred friendship networks based on co-location data.](image)

Fig. 4. Inferred friendship networks based on co-location data. Each graph displays the nodes physically separated by modularity class and each community is represented by a different color. Links represent co-location events detected between two individuals. The others are same with Fig. 3.

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outdegree (density)</td>
<td>Overall tendency to connect to others</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>The tendency to have reciprocal choice of network partners (e.g. if A calls B, B also tends to call A.)</td>
</tr>
<tr>
<td>Transitive triplets</td>
<td>The tendency for the choice to form a triadic closure (e.g. if A calls B, and B calls C, then A also tends to call C.)</td>
</tr>
<tr>
<td>Indegree-related popularity</td>
<td>The tendency for a popular individual (e.g. one who receives more calls than others) to become even more popular (e.g. receiving even more calls)</td>
</tr>
<tr>
<td>Balance</td>
<td>The tendency to connect to others with similar connection choice (e.g. reciprocity choice)</td>
</tr>
<tr>
<td>Covariate alter (e.g. gender alter)</td>
<td>Influence of a covariate on popularity in the network (e.g. gender difference in the degree of popularity)</td>
</tr>
<tr>
<td>Covariate ego (e.g. gender ego)</td>
<td>Influence of covariate on activity in the network (e.g. gender difference in the degree of activity)</td>
</tr>
<tr>
<td>Covariate homophily (e.g. similarity in code-switching frequency)</td>
<td>The tendency for individuals with similar attributes to connect with each other (e.g. individuals with similar code-switching frequency tend to connect with each other.)</td>
</tr>
<tr>
<td>Behavior assimilation</td>
<td>Tendency to assimilate to network partners’ average behavior (e.g. individuals tend to assimilate their code-switching use to match that of their network partners.)</td>
</tr>
</tbody>
</table>

1 Outdegree was used for directional networks such as calls and SMS, and density was used for non-directional networks such as physical co-location.
such that participants who code-switched more also tended to be more active in SMS ($p<0.01$). On the other hand, SMS network appeared to have an effect on code-switching use. Participants tended to assimilate their code-switching use to match that of their SMS partners ($p=0.07$).

As for physical co-location network (see Table VII), we found a significant negative effect in density ($p<0.001$), where there was a tendency for participants to be physically close to selective others. A significant transitive triad effect and a balance effect supported this finding (both $p<0.001$). The former indicated that participants’ physical location with their partners tended toward triadic closure, and the latter indicated that participants tended to be physically close to those who shared similar choice of physical partners. Class, nationality (though only marginally significant) and code-switching use showed a homophily effect on participants’ physical partners, such that participants were physically closer to those who were from the same class and/or of the same nationality, and those who code-switched as often as themselves. We found no effect of physical co-location on code-switching dynamics.

### IV. DISCUSSION

Our present study made use of advanced mobile phone technologies, such as Bluetooth scanning and automated pushing surveys with self-reported questionnaires to examine how mobile phone usage and physical co-location may predict social network and psychological well-being of college students. In this study, we first demonstrated that behavioral data collected by mobile phones could reliably infer real friendship networks, such that physical co-location (detected by Bluetooth scans in mobile phones) correlated strongly with self-reported friendship. This is consistent with the finding of Eagle and colleagues [6].

We further demonstrated that mobile phone usage was significantly related to psychological well-being, such as loneliness, sense of community support and adaptation to college life. Students who received a greater amount of phone calls, especially during weekdays, tended to report lower levels of loneliness, better adjustment to college life, and a greater sense of class community. This implies that phone calls may act as a form of social support for the students, playing a key role in friendship building and development.

In addition, we discovered that while students tended to avoid connecting with others not in their mobile phone network, they tended to form a related circle of call partners. In other words, reciprocity and triadic interaction played an important role in the formation of students’ call network. Although there was no significant effect of students’ SMS network, triadic interaction was not significant in the SMS network. One possible explanation could be due to the differences in nature between phone calls and SMS, that is, students might have used call and SMS for different purposes. SMS is usually used for relaying messages in a quick manner, hence, students may not need to form a triadic SMS network as long as messages are relayed successfully. Call, on the other hand, may be used for more in-depth communication and networking. This may account for the fact that popularity in received calls, but not SMS, has a self-reinforcing effect (the Matthew Effect [23], [24]).

Finally, we found an effect of natural grouping on students’ mobile social network. Students tended to call, SMS, and be physically close to those who were from the same cohort class and same nationality as themselves. In addition, code-switching
behavior was found to have a significant influence on students' social network. Students tended to be around others who code-switched as often as themselves, and those who code-switched more also tended to be more active in SMS. Interestingly, SMS, but not call and co-location, appeared to have an assimilation effect on code-switching use. Students tended to adapt their behavior (in this case, language use) to their social networks as measured by their findings that students tended to adapt their behavior (in this case, message thread. Nevertheless, our results revealed an important effect on code-switching use. Students tended to adapt their behavior (in this case, language use) to their social networks as measured by their mobile phone usage and co-location.

V. CONCLUSION

Our study contributes to the growing literature on the use of mobile phones to study human interactions over an extended period, providing critical insights into the evolution of social networks. Our study also fills an important gap in understanding how human interactions in social networks may in turn, influence our own behavior, specifically language use and psychological wellbeing. Using the latest technology in mobile phone sensing, our study has provided important evidence that social network, mobile phone usage, and language use may co-evolve over time.

REFERENCES