Mining Smartphone Data to Classify Life-Facets of Social Relationships

Jun-Ki Min, Jason Wiese, Jason I. Hong, John Zimmerman
Human-Computer Interaction Institute, Carnegie Mellon University
loomlike@gmail.com, {jwwiese, jasonh, johnz}@cs.cmu.edu

ABSTRACT

People engage with many overlapping social networks and enact diverse social roles across different facets of their lives. Unfortunately, many online social networking services reduce most people's contacts to "friend." A richer computational model of relationships would be useful for a number of applications such as managing privacy settings and organizing communications. In this paper, we take a step towards a richer computational model by using call and SMS logs from mobile phones to classifying contacts according to life facet (*family*, *work*, and *social*). We extract various features such as communication intensity, regularity, medium, and temporal tendency, and classify the relationships using machine-learning techniques. Our experimental results on 40 users showed that we could classify life facets with up to 90% accuracy. The most relevant features include call duration, channel selection, and time of day for the communication.

Author Keywords

Mobile social network; interpersonal relationships mining; life-facets; smartphone.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Algorithms; Experimentation; Human Factors.

INTRODUCTION

People enact many different social roles as they move between contexts and interact with different people. A woman might enact the role of mother, wife, daughter, sister, neighbor, supervisor, colleague, teammate, subordinate, chairperson, coach, and music patron all within a single day. Social roles provide a sort of invisible structure, guiding people in their choice of actions in various social situations. Computational systems, interestingly, often have little or no understanding of the many roles a person might play, of the behavioral expectations associated with these roles, or of the specific role a person is enacting when interacting with a system. Even social devices and services like smartphones and social networking services operate with a tremendously limited understanding of social role, often classifying everyone as "friend" or "friend of friend." Systems with a richer understand of role could assist people in a variety of ways. At the highest level, these systems could engage in much more situationally appropriate behavior. More pragmatically, these systems could help by organizing and prioritizing communications, by working to prevent unwanted self-disclosure or socially inappropriate behavior when sharing on social networking services (SNSs), by reminding people of the role they should be enacting before taking a phone call or engaging in other mediated communication, and by mining situational enactments of different roles to help systems better understand the meaning of a place, of a situation, or of the services people might most desire.

CSCW '13, February 23–27, 2013, San Antonio, Texas, USA.

Some computational systems provide tools for users to manually label groups and assign their contacts to these groups; however, people do not appear to use these tools. One recent study reported that only 16% of people create any contact groups on their mobile phones [20]. In addition, Facebook reported that less than 5% of users create groups within their set of friends [23]. We suspect that these features are rarely used because people do not perceive enough value in improved services to invest the time and attention necessary to categorize the hundreds of contacts they digitally maintain. The challenge is greater than simply classifying each contact once. Research shows relationships are dynamic [35,3]. Kelley et al. argue that groups created for privacy purposes need to be periodically updated because of changes in relationships [25].

Our goal is to develop methods for systems to infer social role at a level of granularity that allows significantly improved service offerings by mining logs of electronic communications and sensor data from smartphones. Data from email and online social networking sites provide rich details on the communications between groups of people. Data from smartphones provide a person's locations as well as proximity to and co-location with others. For example, previous work demonstrated how information on co-location patterns could be used to predict if two people are friends [10]. In addition, phones provide SMS and phone logs, capturing the: who, when, initiator/receiver, and the duration/length of many communication events. The integration of all of this data offers the opportunity to model human behavior and social interactions at a scale and fidelity not previously possible.

This paper documents our investigation of one part of this bigger picture, namely using smartphone communication data (contact list, call logs, and SMS data) to classify the life facet of a user's contact as *family*, *work*, or *social*. Life facets are distinct domains within which people enact different roles. We chose these facets based on previous social networking research by Ozenc and Farnham [36]. While these facets are broad and coarse-grained, they provide initial evidence that concepts such as social role can be learned. Based on communication log data from 40 participants, we used machine-learning techniques to classify life facets with 90% accuracy for contacts users had communicated with.

This paper provides two contributions. First, it provides a set of extracted smartphones features that help to mine life facets. Second, it provides details of models built to classify life facets and an evaluation of these models based on a data set of 16,940 calls and 63,900 SMS messages.

BACKGROUND

Kinds of Relationships

Both social identity theory (from social psych) and identity theory (from sociology) agree that social role is a critical component in understanding human relationships [17,39,40]. People enact many different social roles and form relationships and affiliations with groups based on these roles. The degree that social role structure and granularity vary depends on individuals; however, social constructs like the life facets of *family*, *work*, and *social* are much more universal [33,36]. In exploring her work/family border theory, Clark suggested that the greater the differences across role domains, the less people engage in across-the-border communication [9]. Similarly, Farnham and Churchill found that one approach people use to manage roles and facets is to match them to specific communication mediums, such as only using the phone to communicate with your mom [17].

Relative to offline social contexts (such as face-to-face interaction), people experience new challenges in online contexts, which tend to be more broadcast oriented. This openness, particularly in online SNS, often results in unwanted self-disclosure, where information that is relevant in one life facet but inappropriate in another manages to spillover [20,26,32]. As a result, users often experience difficulty in regulating boundaries [20,26]. Ozenc and Farnham explored how users can leverage bounded, group-based sharing models with online media, in the contexts of *family*, *work*, and *social*, and emphasized the ubiquity of smartphones as an important tool for managing these relationships [36].

Together, this work suggests that the distinctions between *family*, *work*, and *social* are general enough to apply broadly to individuals, and that these distinctions are important and defining factors in people's lives. Furthermore these studies highlight the challenges of maintaining boundaries when using modern technologies. Our work is an effort to help address these concerns by providing a mechanism for systems to gain more detailed information about the specific relationship between people.

Social Network Analysis and Group Mining

Researchers have addressed SN analysis on smartphones and SNS using both supervised and unsupervised approaches [3,6,32]. Many of them have focused on tie strength, defining its properties as four dimensions: amount of time, intimacy, intensity, and reciprocal services [19]. Studies have shown that the vast majority of interaction on SNS is with small numbers of strong ties. For example, recent work by Burke suggested that the average number of friends on Facebook is around 180 [5] (though many users have many more), while most people on Facebook only interact regularly with 4 to 6 people [42]. A different study examined people who posted and tagged pictures of each other on Facebook, and found that on average people had 6.6 such "friends" [8].

Based on network information like tie strength, researchers have tried to analyze distinct groups within a SN. Skeels and Grudin [38] found that people faced many tensions when they tried to manage the co-presence of multiple groups within their network. Lampinen et al. [26] showed that people address group co-presence based on behavioral strategies such as dividing the platform into separate spaces, using suitable channels of communication, and performing self-censorship. Several studies have demonstrated users' desire to create groups of contacts for practical applications like multi-level access control when sharing content [12,25,38]. Olson et al. [34] found that people decide with whom to share information based on the type of relationship, such as family and coworker. For example, Gilbert and Karahalios suggested that privacy controls based on tie strength might help to segment a user's SN into meaningful groups, and achieved 85% accuracy on binary classification of people's strong or weak tie [18]. With their study on people's privacy concern, Jones and O'Neill found six criteria of grouping that people commonly considered: social circles and cliques; tie strength; temporal episodes; geographical locations; functional roles; and organizational boundaries [24].

Meanwhile, as smartphones have become popular and more advanced, they have become an invaluable tool for making inferences about a person's physical and social identities. Reality Mining was one of the first attempts to utilize mobile data with survey studies pertaining to human social behavior [15], showing that a user's social network could be detected by using call records, cellular-tower IDs, and Bluetooth proximity logs. Since then, numerous studies on mobile phones have inferred a broad array of attributes, including: identifying users' demographic information (ethnicity, age, and marital status) [2]; personality [7]; predicting mobile interaction [30]; predicting friendships between people based on co-location patterns [10]; predicting willingness to share different kinds of contextual information [41]; and health-related behaviors that are correlated with social interactions [27]. Relatively little work has attempted to engage the relationship mining problems using data available on mobile phones. The studies that do engage this tend to focus on one type of relationship, namely friendship [35,16], despite literature that indicates that people have overlapping groups of relationships across their personal social networks [12,26,36].

In this paper we focus on mining interpersonal relationships based on the three main life facets (*family*, *work*, and *social*) that appeared most robust and general in the aforementioned literature. We used people's mobile phone data (contact list, call logs and SMS logs) to classify relationship types, and demonstrate which features can be effective in identifying each relationship. As a first step towards testing the feasibility of this data to infer the top-level life facets, we collected historic data available on mobile phones. Our work allows us to examine the relationship between people's communication behaviors and the life facets that their social connections belong to, a topic not previously explored in the literature above.

DATA COLLECTION

Our study ran from January 2012 through April 2012. We recruited participants living in the United States by posting ads in several online bulletin boards and websites. We then gathered data from their Android smartphones and asked them to categorize their relationships with contacts to provide ground truth. To protect our participants' privacy, we anonymized contact names. While we did not collect the content of messages, we did collect descriptive information such as email domain name, first six digits of phone numbers, and postal address without street name, to use for machine learning features.

Participants and Procedure

40 participants (13 male and 27 female) were selected based on several criteria: age ($\geq$ 18), social network membership (on Facebook with at least 50 "friends"), and mobile device usage (have used the same Android phone for at least six months, to have a useful SMS and call log). 55% of our participants were graduate or undergraduate students, 35% were employed, and the rest (10%) were unemployed. The age of the participants ranged from 19 to 50 years (*mean* = 28.0 years, σ = 8.9). Participants were instructed to complete the entire study within two weeks, and were compensated $80 USD.

Participants downloaded our Android app. This app copied their contact list, call log, and SMS log to a database file that they could download to their computers. They also downloaded their friend list from Facebook, which we used to help select which friends they would label. Participants then uploaded these data files to our server.

Next, participants were asked to list the names of people in specific social groups, including immediate and extended family members; people they currently live with; work with; feel close to; and do hobbies with. We selected these labels based on past qualitative work [28,39,41]. We then created a list of 70 contacts for each participant to label with relationship information. We included all the people each participant listed in the previous step. Next, we selected the top 15 most frequently communicated with individuals for phone, SMS, and Facebook (the number of wall posts, comments, and Facebook messages from a contact), respectively. Participants resolved duplicates, including people who had different names in the different communication mediums or that had multiple listings in the contacts list. We then added randomly selected individuals from their phone's contact list and Facebook friend list. Again, participants identified duplicates. This resulted in a list of 70 distinct names for each participant (hereafter referred to as the "*70-contact* list").

We asked participants to provide basic demographic information about each contact from the 70-contact list, such as sex and approximate age (see Figure 1a). Participants were also asked to rate their perceived closeness with each contact on a 5-point scale ranging from *very distant* to *very close* [1,5,11,37] and to indicate how many years they had known the contact. We asked participants how frequently they interact with each contact using a 7-point scale ranging from *less than yearly* to *daily*. They used this scale to describe frequency for face-to-face interaction, overall communication through mobile and online channels, mobile phone call, text messaging, and interaction via Facebook, respectively.

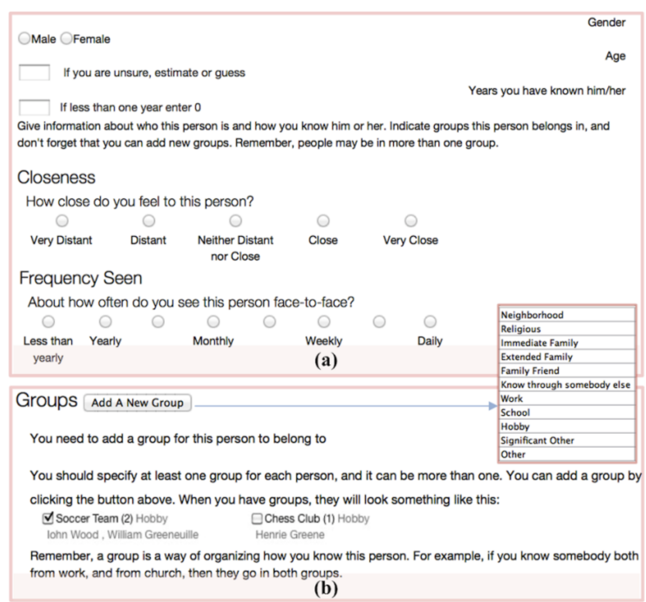

Figure 1. Part of our study webpages. Each participant is asked to (a) answer questions about her relationship with a selected contact; and (b) label what groups this contact belongs to. We used this data as ground truth

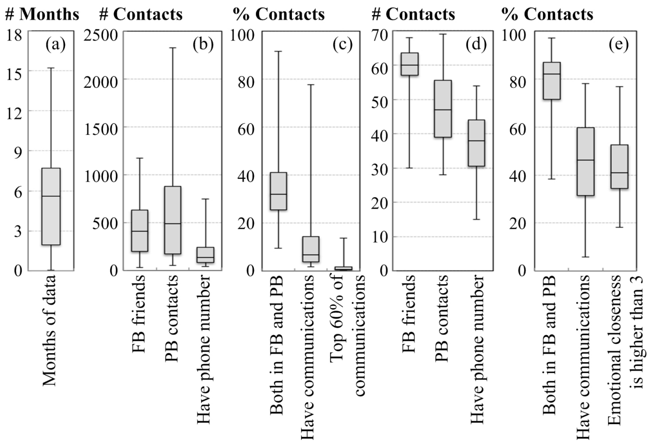

Figure 2. Outline of our data set. (a) Months of data; (b) #contacts in the Facebook (FB) and mobile phonebook (PB) from the entire data set; (c) ratio of contacts for the PB contacts in (b); (d) #contacts in the FB and PB for the *70-contact list*; and (e) ratio of contacts for the PB contacts in (d).

Next, participants assigned each contact to at least one group (we encouraged them to indicate multiple group affiliations when relevant, see Figure 1b). For example, if a participant and her friend went to college together, and they both attend or attended the same church, the participant could place them in two groups. Each group was defined to be one of 11 different categories corresponding to the three life facets: *family* (immediate family, extended family, and significant other), *work*, and *social* (neighborhood, religious, family friend, knowing through somebody else, school, hobby, and others). We selected the names of these sub-categories by combining several sources [22,25,39,41], aiming to strike a balance between comprehensiveness (for richness of data) and simplicity (to prevent errors and minimize burden for participants).

RESULTS

In total we gathered the logs for 24,370 phonebook contacts, 16,940 calls, 63,893 text messages (SMS), and 1,853 multimedia messages (MMS). Note that Android automatically syncs its phonebook contact list with online contact lists such as Outlook, Gmail, and Facebook; the system merges identical contacts, but there are frequently duplicates. We eliminated duplicates by matching names participants provided or by comparing a name and phone number. We ignored calls and messages from/to phone numbers not in a participant's phonebook.

The number of contacts in participants' phonebooks varied: 14 to 2355 contacts (Q1 = 236, Med = 519, Q3 = 962), where 12 to 772 contacts (Q1 = 125, Med = 172, Q3 = 301) had phone numbers in their profile information (the rest of these in- phonebook contacts only had "online" information like email address). Past research found that people have consistent communication with between 7 to 20 people [5,22]. This pattern bore out in our data set. For example, as shown in Figure 2c, the top 60% of calls and messages were made with just 1 to 31 contacts (Q1 = 5, Med = 6, Q3 = 8). Descriptive statistics for the *70-contact list* are shown in Figure 2(d, e); we will focus on this list (2680 contacts; 107 were not in our target facets and 13 had no label, see Table 1 and Figure 3) for the remaining experiments and analysis. For the in-phonebook contacts (1847 contacts; a subset of *70-contact list*), our participants had communicated with 817 contacts through mobile phone call or messaging during their logged period. In addition to the *70-contact list*, we will refer these two subsets as the "*in-phonebook list*" and "*communication list*", respectively. Note that the *70-contact list* includes people who are important to participants (such as live-with, family, coworker, social friends, and most frequently communicated with). Thus, in the *70-contact list*, about half of contacts had a close relationship (self-reported closeness higher than 3 out of 5). Also, more than 80% of the *in-phonebook list* overlapped with the Facebook network as plotted in Figure 2e.

Facet	Group-Category	# Contacts	% Participants		Example labels created by participants
Family	Immediate family	181	90	90	Home, Parents, Close Family, Siblings, Children
	Extended family	219		75	Relatives, Cousins, Uncle, Brother-in-laws, Mother's side family
	Significant other	23		35	Boyfriends, Husband, Ex-boyfriends, Partner, Sig other
Work	Work	305	72.5		Friends of work, Clients, Senior, Previously worked with
Social	School	1136	100	90	UIC, Pitt students, Indiana high school, Roommates this year
	Hobby	78		25	Poker, Marathon, Chess, Old dance people
	Neighborhood	139		40	Current neighbors, Roommate, Met while lived in Morgan park
	Religious	8		7.5	Church friends
	Family friend	100		3.25	Friends of parents, Children's friends' parents
	Know through somebody	260		55	People from Greensburg, Boyfriend's friends, Online friends
	Others [a]	452 + 22	80		Facebook friends, My doctor, Not sure, Mentor

a. We had an "Others" category (4% in our data set) for contacts who did not fit into one of the three facets such as "My doctor" or "Can't remember this person." Some participants used this to distinguish "best friends" from social friends. We used the self-reported closeness to assign the contacts in the Others category into the *social* facet (closeness ≥ 2, 452 contacts; or have another *social*-type label, 22 contacts), and rejected the rest 107 contacts.

Table 1. Relationship categories and examples of group labels created by our participants. Column "# Contacts" shows the number of contacts assigned into each category label (out of a total 2680 contacts from 40 participants, allowing multiple labels), and "%Participants" shows the percentage of participants that labeled at least one contact in a given category.

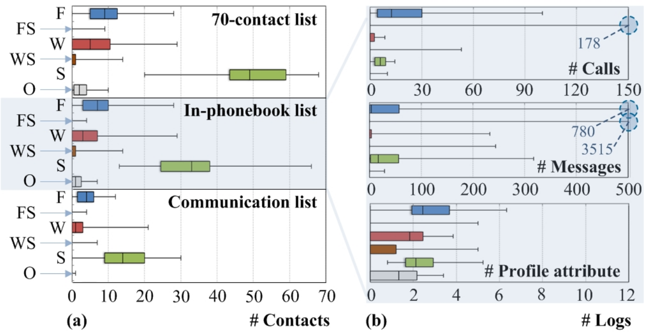

Figure 3. Distribution of *family* (F), *work* (W), *social* (S), and the others (O) contacts (FS and WS denote contacts in multiple facets; we rejected O contacts): (a) The number of contacts in each list; and (b) the log size in the *in-phonebook list*

Relationship Categories

We used three life facets (*family*, *work*, and *social*) rather than using all the group categories (see Table 1 and Figure 3). Note that if a contact was assigned to two or more groups crossing different faceted boundaries, he/she was labeled with multiple classes. In our data set, 14% of contacts were *family* (here we included "significant other"). Since more than half of our participants were students, we have a relatively small number of *work* type contacts (11%). School presents a difficult challenge for our analysis. Some student-participants labeled part of their relationships as *work*, while others reported all of their school contacts as *social*. This demonstrates a limitation of simplifying the groups down to these three facets.

We found that *social* is the most common life facet. In our data set, 70% of contacts were *social*, with diverse group labels like schoolmates, poker friends, church friends, and online friends. Most of the cross-labeled contacts were also within *social* categories, such as school friends who play chess together (school and hobby) and roommates (school and neighborhood). Only 0.9% and 2.3% of contacts were in multiple facets of *family-social* and *work-social*, respectively (see Figure 3a).

APPLYING MACHINE LEARNING TO LIFE FACETS

Mobile Communication Patterns

Based on the observations of our data set and several sources of references, we defined five factors that characterize communication patterns in the context of the life facets: intensity, regularity, temporal tendency, channel selection/avoidance, and maintenance cost.

Intensity and regularity: The number of and duration of communication actions has been used to predict tie strength in past work [22,37]. For example, as illustrated in Figure 3(b), voice calls were very common for *family* relationships. Meanwhile, office interactions and personal interactions show different regularity [13]. Figure 4 shows examples of communication patterns between a contact and a participant, where *work*-type interaction (Figure 4b) was less regular than those of *family* and *social* (Figure 4a and 4c respectively).

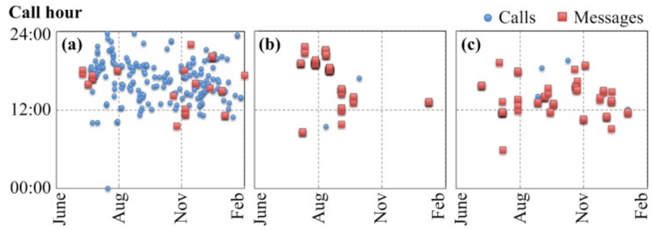

Figure 4. Examples of communication patterns between a participant and a contact: (a) *family*, (b) *work*, and (c) *social*

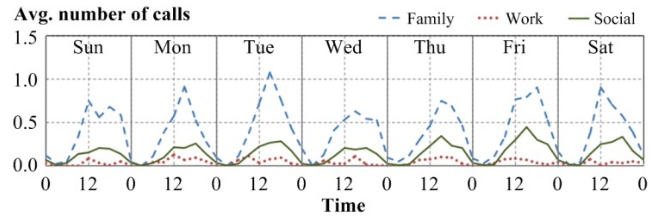

Figure 5. Averaged number of calls over the contacts in the *communication list*. Users preferred to call *social* contacts after work-hour while less with *work* facet contacts on the weekends.

Temporal tendency: In their friends-acquaintances work, Eagle and Pentland observed the temporal tendency in contacting people [15]. Our participants called with their *family*, *work*, and *social* contacts at different times of day and days of the week (see Figure 5). For example, people frequently called their *family* members in the afternoon (12~15h) and evenings (15~18h), and contacted *social* friends more on Friday afternoon. They were less likely to call *work* facet contacts on the weekend.

Channel selection and avoidance: People tend to favor a certain communication medium based on the person they are communicating with [29]. According to our survey result, participants communicated via face-to-face, calls, text messages, and Facebook with different priorities based on the relationship of the contact.

Maintenance cost: Roberts and Dunbar [37] suggested that people apply different amounts of effort in maintaining different kinds of relationships, where effort (cost) can be evaluated based on the time to last contact between two individuals. However, this value could be smaller than the actual cost by chance according to the data collection date. To minimize this coincidence, we measure the number of communications in the past two weeks (short-term view) and in the past three months (relatively long-term view) as the maintenance cost of the given relationship. In the following section, we describe how we operationalized these categories into specific machine learning features.

Feature Extraction

We defined 153 features from mobile data: 17 from phonebook, 66 from call logs, 36 from message logs, and 34 from communication (call + message) logs (see Table 2).

Phonebook contact-profile features (profile features): We measured the similarity between our participants and their contacts in terms of last name, email domain, first six digits of phone number (area code + exchange), and company name. For email domain, we excluded several popular free-hosting domains, such as @gmail.com. We also captured the labels on each contact's phone numbers, such as *home*, *work, mobile,* and *other*, which are pre-defined in Android. Since only few participants had their own and contacts' postal-address in the phonebook, we simply checked if the contact's postal address exists, and measure a binary feature for each label (*home*, *work*, and *other*) of the address. We also had additional binary features for additional contact information, such as ringtones, notes, or if the contact was marked as a favorite in the Android contact list.

Finally, we calculated the completion of each profile. This serves as a weak proxy to measure the user's effort in filling in the contact information. Android phones can automatically fill contact list data by syncing with online SNS such as Google+ and Facebook. In our *70-contact list* data set, 50% of mobile contacts had photos, suggesting that about a half of participants used auto-sync. Interestingly, we still found that the majority of contacts had less than 3 profile attributes, including name, a phone number, and an email address (see Figure 3b). Thus, more contact information could still be a potentially useful signal.

Call features: We created several features to examine the *intensity* of voice communications, including total number of and duration of calls for each contact. We counted the number of lengthy calls, in which their duration is twice the average duration. To measure the *regularity* of calls, we calculated five statistical values (AVG, STD, MIN, MAX, and MED) for the number of calls per week. For the duration of calls, we calculated AVG and MAX to measure a simple regularity and burst-ness, respectively. We also measured users' *channel selection and avoidance* behaviors by calculating the ratio of outgoing calls, failed (duration < 5 seconds) calls, and missed calls. *Temporal tendency* and *maintenance cost* were also extracted from call logs as shown in Table 2.

Message (MSG) features: We defined many message features similar to the call features as shown in Table 2. Here, we counted the length of MMS as 160 characters, which is the maximum length of SMS. With their large-scale text messaging study, Battestini et al. applied a time metric of a 20-minute response time to group messages together as a conversation [4]. We used this metric to capture the ratio of replied messages.

Communication (COMM) features: Messages frequently involve other communication channels [4]. For example, people often call back using voice in response to an SMS. Therefore, we considered calls and messages equally as the COMM features. For the intensity, regularity, and channel selection features, we calculated *total #COMMs*, the *#days had COMMs / days logged*, and the ratio of calls (*#calls / total COMMs*), respectively. We also had features checking *COMMs* on four days that, in the United States, often carry some kind of meaning. These days included Thanksgiving, Christmas, New Year's Day, and Valentines' Day.

Finally, we extracted four features from survey results, such as same gender, age difference, is-Facebook friend, and frequency seen, referring to them as the *survey features* (see Table 2). Current smartphones do not have these data, but

Modality		Variables	Description
Survey		Same gender; age difference; is Facebook friend; frequency seen	Available from SNS and GPS
Profile features		Similarity of {last name, email domain, company name, phone number}; has {photo, ringtone, webpage, note, affiliation, work-phone & address, home-phone & address, email}; is starred; event type; completion of profile	Contact information
Mobile-communication features	Call logs	Total {#, dur.} calls; total #lengthy-calls;	Intensity
		{AVG, STD, MIN, MAX, MED} #{outgoing, incoming} calls per week; #days called / days logged; {AVG, MAX} dur. {outgoing, incoming} calls	Regularity
		{#, dur.} outgoing calls / total calls; #{failed, missed} calls / outgoing (incoming) calls	Ch. selection & avoidance
		{#, dur.} calls at {times of a day, days of a week} / total calls; #{lengthy, failed, missed} calls at {times of a day, days of a week} / total (outgoing, incoming) calls	Temporal tendency
		{#, dur.} calls for the {past two weeks, past three months} / total calls	Maintenance cost
	MSG logs	Total {#, length} MSGs	Intensity
		{AVG, STD, MIN, MAX, MED} #{outgoing, incoming} MSGs per week; #days MSGed / days logged;	Regularity
		{#, length} outgoing MSGs / total MSGs; #replied MSGs / total MSGs	Ch. selection & avoidance
		{#, length} MSGs at {times of a day, days of a week} / total MSGs	Temporal tendency
		{#, length} MSGs for the {past two weeks, past three months} / total MSGs	Maintenance cost
	COMM (Call & MSG) logs	Total #COMMs	Intensity
		#Days had COMMs / days logged	Regularity
		#Outgoing COMMs / total COMMs; #calls / total COMMs	Ch. selection
		# COMMs at {times of a day, days of a week} / total COMMs; #calls / total COMMs at {times of a day, days of a week}; outgoing COMMs at {times of a day, days of a week, holidays} / total COMMs	Temporal tendency
		# COMMs for the {past two weeks, past three months} / total COMMs	Maintenance cost

Table 2. Extracted 153 features from mobile data. We used four time segments and four day of week segments: morning (6AM ~ noon), afternoon (noon ~ 6PM), evening (6PM ~ midnight), and night (midnight ~ 6AM); Weekdays (Mon ~ Thu), Fri, Sat, and Sun. Note that "total" communication denotes all previous (logged) communications with the contact.

we included these as features since some of the information could be obtained from today's SNS, and others could be captured by future smartphones. For example, demographic information such as gender and age can be automatically retrieved from Facebook profiles (or estimated from smartphone usage patterns [2]), and face-to-face communication could be detected by using Bluetooth proximity [13] or GPS co-location [10].

Life Facet Classification

We evaluated how well we could classify the life facets with different sets of features across different data sets. As noted by a previous study on friends and non-friends [14], the size of different life facets are not necessarily balanced. In our data set, two-thirds of contacts were *social*, meaning that a null classifier could achieve good accuracy just by guessing "*social.*" Therefore, when evaluating a given data set, we randomly sampled contacts to balance their classes (non-replacement sampling with WEKA toolkit [21]). For the classification test, we conducted three runs of 10-fold cross-validation, where classes were re-sampled for each run. Table 3 shows the class distributions of the original datasets and resampled datasets.

We first analyzed associations of the extracted features to the facets by calculating Pearson correlations (see Figure 6). Here, the *communication list* data set is used to highlight people's mobile communication behaviors over different facets. As shown in Figure 6(a), the *frequency seen* feature was related to the *work* facet. The age difference between a user and contact was more correlated with the *social* facet than the others, which implies that people hang out more with similar age contacts. For the profile features, *Is-starred* ("favorite" list in the Android phonebook) and *Has-photo* features were strongly associated with *family* and *social* types respectively; photos in phonebooks are mostly synced from Facebook. *Event type* (*birthday*, *anniversary*, and *other*) was also highly correlated with *family* and *social*.

Intensity and regularity communication features showed positive scores for the *family* and *social* relationships, respectively; while *work* type was negatively related to those features. Based on the correlation values from the *#Calls in evening* and *weekends*, we also confirmed that people prefer not to have calls with *work* contacts after work-hours and on weekends.

To evaluate the life facet classification, we built a model with SVMs (Polynomial kernel; d = 2). Since an SVM is

Type of data set	Original data sets			Evaluation data sets		
	Family	*Work*	*Social*	*Family*	*Work*	*Social*
70-contact list	419	305	1956	501	521	537
In-phonebook list	315	227	1305	399	422	400
Communication list	168	95	554	202	198	210

Table 3. The number of contacts for the original data set and the evaluation set (sum of three runs). Note that we conducted three runs of 10-fold cross-validation, where classes were balanced by using non-replacement sampling for each run.

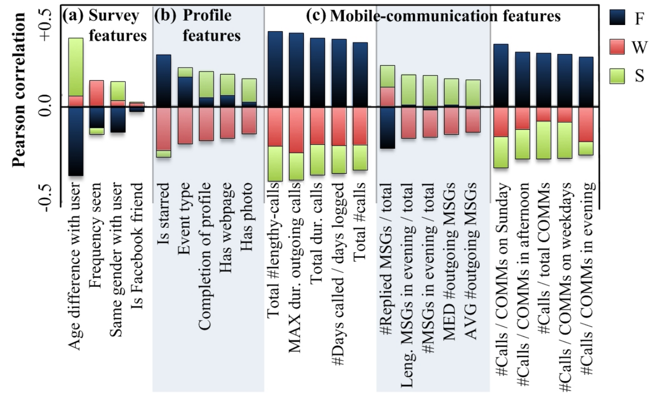

Figure 6. The most correlated (Pearson correlation) features of each modality with each life facet. Call and message features were positively correlated with *family* and *social*, respectively.

Data set	Method	Survey features	Profile features	Mobile comm. features	Profile+ mobile comm.	Using all
70-contact list	C4.5	68.0(5.1)	NA	NA	NA	74.2(4.4)
	NB	72.4(5.6)	NA	NA	NA	67.1(3.6)
	SVM	65.5(4.4)	NA	NA	NA	81.0(4.5)
In-phone book list	C4.5	69.1(6.9)	52.8(7.7)	64.7(6.0)	64.5(7.8)	76.1(7.5)
	NB	72.5(7.2)	54.3(6.6)	65.5(5.5)	65.4(5.6)	65.6(5.5)
	SVM	66.7(7.4)	**51.1**(7.2)	**67.9**(6.4)	**68.5**(6.0)	**83.1**(5.9)
Comm. list	C4.5	66.0(9.9)	53.1(8.7)	75.8(7.2)	75.9(6.7)	75.6(5.5)
	NB	69.1(10.)	55.2(9.0)	76.4(8.8)	76.4(8.8)	76.6(8.6)
	SVM	60.8(10.)	52.9(8.4)	87.1(5.0)	88.0(5.3)	90.5(4.8)

Table 4. Ave. accuracy (and Std.) of three runs of ten-fold cross-validation for predicting life facets. Note that online friends in the *70-contact list* have no mobile features (shown as NA in the table). Bold face denotes the results shown in Table 5.

originally a binary classifier, we composed 3 pairwise SVMs (*family*-vs.-*work*, *family*-vs.-*social*, and *work*-vs.-*social*) and combined their outputs with a maximum voting method. We also compared the result with two different models using WEKA [21]: rule-based model (decision tree C4.5) and probabilistic model (naïve Bayes; NB). All the models were tested on three types of evaluation data sets shown in Table 3.

For the *70-contact list* data set, the SVMs showed 65.5% accuracy with survey features and increased to 81.0% when using all features (see Table 4). Note that half of the contacts in the *70-contact list* were online friends, that is, the profile and mobile-communication features did not contain their data. Nevertheless, this approach achieves a 13.2% improvement in accuracy in using survey feature. The SVMs yielded 68.5% accuracy on the *in-phonebook list* using the profile and mobile-communication features together. Accuracy went up to 88% for contacts that users actually communicated with (the *communication list* set).

Using profile features alone, all the classifiers showed lower accuracies since users did not have enough contact information in their phonebook (see Figure 3b). For example, on the *in-phonebook list*, SVMs yielded only

51.1%. Table 5a shows the confusion matrix of the result. There were 396 total *family* members evaluated, where 175 were correctly classified as *family*, while 186 and 35 were misclassified as *work* and *social*, respectively. Interestingly, the profile features were effective in classifying *work*-type contacts but were not so for other facets. This is because the profile variables that strongly correlated with the *family* facet, like *last name* and *Is-starred*, were missing on many contacts or only associated with a few strongest ties. On the other hand, we could classify *family* and *social* contacts correctly by using mobile communication features while many of the *work*-type contacts were confused as *social* because of no communication logs (see Table 5b).

The combination of profile and communication features failed to classify the *work* facet, shown in Table 5(b, c). We have two possible explanations. First, the number of the profile features (17) was much smaller than the number of the communication features (136), and thus might be less influential within the combined feature vector. Second, the profile features had many missing values and SVM does not handle missing data well. To address these issues, we need a feature selection scheme or classifier combination approach with two different models built on each type of features. We defer this line of investigation to future work.

(a)	*Family*	*Work*	*Social*
Family	175	186	35
Work	68	247	71
Social	77	160	202

(b)	*Family*	*Work*	*Social*
Family	288	31	76
Work	49	111	198
Social	15	23	430

(c)	*Family*	*Work*	*Social*
Family	290	28	76
Work	46	114	198
Social	15	22	432

(d)	*Family*	*Work*	*Social*
Family	298	43	53
Work	27	305	44
Social	20	20	411

Table 5. Confusion matrices (row: actual, column: predicted) of SVMs for the *in-phonebook list* with (a) profile features, (b) mobile-communication, (c) both profile and communication, and (d) using all (including survey features). Row sum of each matrix could be different because of multi-labeled samples.

	Feature name	Info. gain (corr.)	Feature values and sample distribution (target : rest facets)
F	Total dur. calls	0.547 (+)	>588 (0.90:0.11)
	Total #lengthy-calls	0.481 (+)	>2 (0.73:0.04)
	#Calls on Sun. / total COMMs	0.478 (+)	>0.02 (0.83:0.10)
	Dur. calls on Sun. / total calls	0.470 (+)	>0 (0.83/0.11)
	#Calls on Sun. / total calls	0.470 (+)	>0 (0.83/0.11)
W	Total dur. calls	0.225 (-)	>16 & ≤588 (0.65:0.17)
	Total #calls	0.217 (-)	>1 & ≤4 (0.52:0.08)
	Dur. calls on weekdays / total calls	0.144 (+)	>0.20 (0.48:0.11)
	MAX dur. incoming calls	0.140 (-)	>16 & ≤574 (0.65:0.26)
	#Calls on weekdays / total calls	0.140 (+)	>0.2 (0.40:0.08)
S	Total dur. calls	0.442 (-)	≤16 (0.72:0.09)
	#Days called / days logged	0.441 (-)	≤0.01 (0.75:0.08)
	Total #calls	0.425 (-)	≤0 (0.69:0.08)
	#Calls / total COMMs	0.421 (-)	≤0.06 (0.80:0.13)
	(#Calls / COMMs) on weekday	0.396 (-)	≤ 0.02 (0.85:0.16)

Table 6. Top five information gain (with ranker [21]) features to predict each facet (for *communication list* data set; all features except the survey feature). Examples of feature values depict people's common contact behaviors.

As shown in Tables 5(c, d), four features from the survey reports (same gender, age difference, face-to-face communication frequency, and is-Facebook friend) resolved confusions on *work*-type contacts. Since the communication features had continuous values, SVMs performed better than the other algorithms. Meanwhile, C4.5 and NB showed higher accuracies on the survey and profile features which consist of discrete values. For the combination of features, however, the SVMs performed much better than the others.

Table 6 shows the top five information gain features for each life facet, where the signs denote positive or negative correlations. For example, *Total duration of calls* was the strongest feature to recognize the *family* facet, where the participants had longer calls (duration > 588 sec.) with their *family*-type contacts (90% of *family* contacts) and not with *work* and *social* contacts (11% of them), denoted as (0.90 : 0.11) in the table. *Family* facet was positively associated with the lengthy-call features and Sunday-calls as well.

For the *work*-type relationships, the call intensity features (number and duration of calls) were negatively correlated and the ratio of *weekday calls* over *total calls* was positively associated with. Participants communicated with their *social* contacts via text-messages for weekdays (negative *#Calls / COMMs on weekday*). *#Days messaged* and *#Calls on Sunday* were also a negatively associated feature to the *work* facet (which are not shown in Table 6), while after-work communications (positive *#Outgoing COMMs in evening* and negative *#Calls in morning*) and messages (positive *Total length of MSGs* and *#Replied MSGs*; less intensive calls) were the characteristics of *social*-type communications for our participants.

The feature values listed in Table 6 might be affected by several factors such as a logging period, user's mobile plans (e.g. free SMS), and personality. However, the table still depicts how people communicate with different facets, and shows these patterns as concrete examples.

DISCUSSION AND FUTURE WORK

We achieved 90.5% accuracy for classifying contacts that the participant communicated with, and 83.1% for classifying all contacts in the participant's phonebook list. For each participant (by leave-one-contact-out cross validation), our model yielded 91.2% accuracy (Max = 100%, Min = 50%, and σ = 11.6), which shows its effectiveness across the different participants except for some uncommon cases. An open question is how good a system would need to be in order to effectively enhance different kinds of applications.

In our test results, errors were mainly due to a lack of information, such as minimal profile information or no call or message logs. For example, we achieved a 14.6% improvement when we used survey features along with the phone features for the *in-phonebook list* data set. However, another open question here is how important it is to classify these individuals. On the one hand, no information suggests that the contact may not be important to the participant. On the other hand, the availability of alternative forms of communication—such as email, iMessage, Skype, Facebook chat, and more—means that these data are not easily available. For example, in our data set, people did not have mobile communications with over half of their phonebook contacts, including 38% (average) of their close contacts (see Figure 2). Both of these questions are issues we plan on exploring in future work.

One limitation of our work is the maximum size of logs stored. By default, Android stores only 500 recent calls and 200 messages per contact. The logs we gathered contained roughly 100 days of communications on average. Another limitation of the results is a relatively small sample size (n = 40). Since our work is a preliminary study with a limited group of subjects who live in US (culture) and have been using Smartphone and Facebook (social medium), we could not say that our results reflect general patterns of interaction that people have (we did not present generalizability of our results because of the scope of this work). There also exists a possible self-selection bias, in that potential participants concerned about privacy might not participate in our study even with our privacy protection. Nevertheless, our results suggest that there is potential for automated tools that can help classify relationship types.

Certain aspects of a relationship such as the duration and history are difficult to measure using this data driven approach. In this paper, we explored how far we can go with relatively straightforward methods, saw where these methods fall short, and then saw what other kinds of data and features can help improve the models. For instance, incorporating location data into our models could give a better sense as to whom a person spend time with, and how much "effort" a person has put into meeting someone. As a concrete example, one idea would be to put more weight on the people contacted when a person goes to a different city. Cranshaw *et al.* showed that co-location features could be used to infer friendship between individuals [10]. Mok *et al.* found that frequency of both face and non-face communication declines with increasing distance [31]. There may be additional correlations between this behavioral pattern and the type of relationships.

Lastly, we focused on three common life facets, but as shown in Table 1, there are a variety of relationships within each facet. Misalignment between different life facets could cause many problems like inappropriate sharing, but fusing different sub-facets into one higher-order class could also be unsuitable. For example, both co-workers and clients are in the *work* facet while the content of relationships are totally different in terms of the kinds of interaction that take place and information sharing concerns to a user.

We also speculate that it may be more effective overall to have five facets, with the two new facets being school and maintenance. In our study, we saw student participants struggle with classifying people as *social* or *work*, and having a specific facet for this purpose could mitigate this problem. Adding a maintenance facet would capture contacts that do not match the other facets such as auto repair, doctor, and pizza shop. These kinds of contacts were excluded in our study, but were something that some participants commented on.

CONCLUSION

People have distinct domains within which they enact different roles, called life facets, and *family*, *work*, and *social* are people's common facets. In this paper, we defined a set of smartphones features to help classify the life facet of a contact, using profile information as well as five categories of communication behaviors, including *intensity*, *regularity*, *temporal tendency*, *channel selection & avoidance*, and *maintenance cost*.

Our models were evaluated with 40 participants. For contacts that the participant had any communication with in their smartphone logs, we could classify those contacts with 90% accuracy. For *family*, call intensity was the strongest factor. For *work*, call intensity (negatively correlated) and weekday communication behaviors were the strongest factors. For *social*, channel selection (dominantly use messages) and communications on weekends were the most relevant features. We also found that using different modalities of features (same gender, age difference, is Facebook friend, and frequency seen) with the mobile communication logs could improve accuracy by 14.6% in classifying all contacts in the participant's phonebook list.

ACKNOWLEDGMENTS

This work is sponsored in part by DARPA I2O Program: DCAPS, ARPA Order No. 9464-00, Program Code: CHEL5, issued by SPAWAR System Center-Pacific under Contract No. N66001-12-C-4196. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the DARPA of the U.S. Government. The first author was also supported by the NRF Korea (NRF-2011-357-D00210).

REFERENCES

1. Ackerman, J.M., Kenrick, D.T., and Schaller, M. Is friendship akin to kinship? *Evolution and Human Behavior*, 28 (2007), 365-374.
2. Altshuler, Y., Fire, M., Aharony, N., Elovici, Y., and Pentland, A. How many makes a crowd? On the evolution of learning as a factor of community coverage. In *Proc. Int. Conf. on Social Computing, Behavioral-Cultural Modeling, and Prediction* (2012), 43-52.
3. Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X.

Group formation in large social networks: membership, growth, and evolution. In *Proc. KDD '06* (2006), 44-54.

4. Battestini, A., Setlur, V., and Sohn, T. A large scale study of text-messaging use. In *Proc. MobileHCI '10* (2010), 229-238.
5. Burke, M. *Reading, Writing, Relationships: The Impact of Social Network Sites on Relationships and Well-Being*. Carnegie Mellon University, Pittsburgh, Pennsylvania, 2011.
6. Chau, M. and Xu, J. Mining communities and their relationships in blogs: A study of online hate groups. *Int. J. of Human-Computer Studies*, 65 (2007), 57-70.
7. Chittaranjan, G., Blom, J., and Gatica-Perez, D. Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, (in Press).
8. Christakis, N.A. and Fowler, J.H. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Little, Brown and Company, New York, 2009.
9. Clark, S.C. Work/family border theory: A new theory of work/family balance. *Human Relations*, 53, (2000), 747-770.
10. Cranshaw, J., Toch, E., Hong, J., Kittur, A., and Sadeh, N. Bridging the gap between physical location and online social networks. In *Proc. Ubicomp '10* (2010), 119-128.
11. Cummings, J.N., Lee, J.B., and Kraut, R. *Communication technology and friendship during the transition from high school to college*. New York: Oxford University Press, City, 2006.
12. DiMicco, J.M. and Millen, D.R. Identity management: Multiple presentations of self in Facebook. In *Proc. GROUP '07* (2007), 383-386.
13. Do, T.M.T. and Gatica-Perez, D. Human interaction discovery in smartphone proximity networks. *Personal and Ubiquitous Computing*, (in Press).
14. Dong, W., Lepri, B., and Pentland, A. Modeling the co-evolution of behaviors and social relationships using mobile phone data. In *Proc. MUM '11* (2011), 134-143.
15. Eagle, N. and Pentland, A. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, 10 (2006), 255-268.
16. Eagle, N., Pentland, A., and Lazer, D. Inferring social network structure using mobile phone data. *Proc. National Academy of Sciences*, 106, 36 (2009), 15274-15278.
17. Farnham, S. and Churchill, E.F. Faceted identity, faceted lives: Social and technical issues with being yourself online. In *Proc. CSCW 2011* (2011), 359-368.
18. Gilbert, E. and Karahalios, K. Predicting tie strength with social media. In *Proc. Int. Conf. on Human Factors in Computing Systems* (2009), 211-220.
19. Granovetter, M.S. The strength of weak ties. *The American Journal of Sociology*, 78, 6 (1973), 1360-1380.
20. Grob, R., Kuhn, M., Wattenhofer, R., and Wirz, M. Cluestr: mobile social networking for enhanced group communication. In. *Proc. Group '09* (2009), 81-90.
21. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. The WEKA data mining software: An update. *SIGKDD Explorations*, 11, 1 (2009), 10-18.
22. Hill, R.A. and Dunbar, R.I.M. Social network size in humans. *Human Nature*, 14 (2003), 53-72.
23. Http://www.cbsnews.com/8301-501465_162-20105561-501465.html
24. Jones, S. and O'Neill, E. Feasibility of structural network clustering for group-based privacy control in social networks. In *Proc. SOUPS* (2010), 1-13.
25. Kelley, P.G., Brewer, R., Mayer, Y., Cranor, L.F., and Sadeh, N.M. An Investigation into Facebook Friend Grouping. In *Proc. INTERACT* (2011), 216-233.
26. Lampinen, A., Tamminen, S., and Oulasvirta, A. All my people right here, right now: Management of group co-presence on a social networking site. In *Proc. GROUP '09* (2009), 281- 290.
27. Madan, A., Cebrian, M., Lazer, D., and Pentland, A. Social sensing for epidemiological behavior change. In *Proc. Ubicomp '10* (2010), 291-300.
28. McCarty, C. Structure in personal networks. *J. of Social Structure*, 3 (2005).
29. Mesch, G.S. Social context and communication channels choice among adolescents. *Computers in Human Behavior*, 25 (2009), 244-251.
30. Min, J.-K. and Cho, S.-B. Mobile human network management and recommendation by probabilistic social mining. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 41, 3 (2011), 761-771.
31. Mok, D., Wellman, B., and Basu, R. Did distance matter before the Internet? Interpersonal contact and support in the 1970s. *Social Networks*, 29 (2007), 430-461.
32. Newman, M.E.J. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38 (2004), 321-330.
33. Nippert-Eng, C.E. *Home and Work*. The University of Chicago Press, 1995.
34. Olson, J., Grudin, J., and Horvitz, E. A study of preferences for sharing and privacy. In *Proc. CHI '05* (2005), 1985-1988.
35. Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K.,Kertész, J., and Barabási, A.-L. Structure and tie strengths in mobile communication networks. *Proc. National Academy of Sciences*, 104 (2007), 7332-7336.
36. Ozenc, F.K. and Farnham, S.D. Life "modes" in social media. In *Proc. CHI '11* (2011), 561-570.
37. Roberts, S.G.B. and Dunbar, R.I.M. Communication in social networks: Effects of kinship, network size, and emotional closeness. *Personal Relationships*, 18 (2011), 439-452.
38. Skeels, M.M. and Grudin, J. When social networks cross boundaries: A case study of workplace use of Facebook and LinkedIn. In *Proc. GROUP '09* (2009), 95-104.
39. Spencer, L. and Pahl, R.E. *Rethinking Friendship: Hidden Solidarities Today*. Princeton University Press, 2006.
40. Tajfel, H. *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*. London: Academic Press, 1978.
41. Wiese, J., Kelley, P.G., Cranor, L.F., Dabbish, L., Hong, J.I., and Zimmerman, J. Are you close with me? Are you nearby? Investigating social groups, closeness, and willingness to share. In *Proc. UbiComp '11* (2011), 197-206.
42. Wilson, C., Boe, B., Sala, A., Puttaswamy, K.P.N., and Zhao, B.Y. User interactions in social networks and their implications. In *Proc. the 4th ACM European Conference on Computer Systems* (2009), 205-218.