Assessing Call and SMS Logs as an Indication of Tie Strength

Jason Wiese, Jun-Ki Min, Jason I. Hong and John Zimmerman

May 2014
CMU-HCII-14-101

Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

How effective are call and SMS logs in modeling tie strength? Frequency and duration of communication has long been cited as a major aspect of tie strength. Intuitively, this makes sense: people communicate with those that they feel close to. Many highly cited research papers have pushed this idea further, using communication as a direct proxy for tie strength. However, this operationalization has not been validated. Our work evaluates this assumption. We collected call and SMS logs and ground truth relationship data from 40 participants. Consistent with theory, we found that frequent or long-duration communication likely indicates a strong tie. However, little or no communication does not necessarily indicate a weak tie. Follow-up interviews revealed several explanations and indicate fundamental challenges for inferring tie strength from communication logs.

1 Introduction

The meteoric adoption of smart phones and social networking services makes it easier to monitor aspects of people's social behavior. Logged communication data collected by smartphones along with traces of online communication services offer a potentially rich way for computer systems to gain more The social sophistication and to better understand the changing interpersonal relationships that people have with each other. Recent work points to benefits from more social sophistication, including digital photo frames that adapt based on the relationships of the people in a place [23]; virtual possession collections that automatically surface based on the social context [34]; and even services that automate the management of privacy settings based on the influence of relationship in people's willingness to share [45].

A wealth of work in social psychology has found that the relationship strength between two people influences their pattern of communication. When a relationship has stronger tie strength, there is generally more communication and when a relationship has weaker tie strength there is generally less communication [18,25]. Subsequent research has operationalized this theory by using phone call frequency as a proxy for tie strength [7,30,35,42].

The social psychology theory on communication frequency and tie strength is based on communication across long periods of time and across all possible channels, including face-to-face. Recent work, however, has operationalized tie strength using data from one or two communication channels and often using small time windows when compared to a person's lifetime. The idea that you can model tie strength with more sparse data is tantalizing but dangerous. If it works, it is a discount approach that can easily be adopted by many applications. However, if sparser communication data does not accurately infer tie strength, then our community's adoption of this untested assumption might lead to many kinds of errors.

To investigate if tie strength can be inferred from sparser communication data sets, we conducted a study on inferring tie strength based on the call and SMS logs stored on smartphones that had been in use for at least six months. We gathered data from 40 participants, including contact lists, call and SMS logs, and a list of friends from Facebook. We also collected ground truth: participants labeled the type of relationship and rated the tie strength of a subset of their contacts and friends.

In line with previous findings conducted using Facebook data, our binary classifier achieved 91.6% accuracy for inferring tie strength. We confirmed that frequent communication indicates strong ties. Interestingly, though, we also observed that *a lack of communication does not necessarily indicate a weak tie*.

To better understand why infrequent communication is sometimes associated with a strong tie, we interviewed a selected set of our participants. We found three explanations: 1) people use many different communication channels, and phone and SMS logs are not representative of

their overall communication patterns; 2) people often feel close with people they frequently see face-to-face, a type of communication not easily observed; and 3) people feel a lingering sense of closeness to friends from a previous stage in their life, but they do not necessarily communicate with these people very often.

Our findings indicate that the use of a few communication channels over a short time window to infer tie strength may not produce reliable results. Depending on the needs of an individual application, systems may need to include many more communication channels, capture face-to-face communications, and track behaviors over several life stages to perform at an appropriate level of confidence.

2 Related Work

2.1 Social Science Research in Social Networks

A large amount of social science research has focused on understanding aspects of human relationships. We highlight some related areas and describe how we operationalize these theories in a machine-learning system.

Numerous studies have examined how many close ties a person has. A study of 3,000 Americans showed that people average four strong ties, with most having between two and six [6]. Another study of 1,000 adults found that most people had 10 friends they meet or speak with weekly [6]. Our work is focused less on counting ties and instead looks to characterize these ties.

A great deal of literature has looked at factors influencing tie strength. Roberts and Dunbar examined how emotional closeness and kinship influence the size of social networks and communication patterns [38,39]. Some studies have used proximity as a proxy for quantity of social interaction between pairs [15,33], showing that communication frequency drops exponentially as members of a pair move farther apart [2,48].

Granovetter identified four dimensions of tie strength: duration, intimacy, intensity, and reciprocity [19]. Gilbert and Karahalios expanded on these dimensions for online communications on Facebook, and built models that could achieve 85% accuracy on binary classification of a person's contacts as strong or weak ties [16]. Using survey data and Facebook data (including passive consumption) from 11,000 participants, Burke developed a linear model for estimating tie strength of contacts on Facebook, which was able to distinguish the closest relationships from others with an accuracy of 71.2% [4]. Our work is similar to the above, and our binary classification accuracy of 91.6% is in line with their results; however, we use a different data source, looking at smartphone usage and not Facebook usage.

Beyond using different data sources, we use a different method for eliciting a participant's contacts to be examined in this study, including: 1) Explicitly asking the participant for their close contacts regardless of whether or not they are in the participant's contact list; and 2) Selecting contacts with whom the participant has frequent communication. By employing this

elicitation approach, we are able to better understand the aggregate of the participant's social relationships. This also enables us to understand the limits of elicitation methods that use only Facebook, or only the mobile contact list.

2.2 Using Sensors to Model High Level Context

An emerging thread in UbiComp research has been the use of sensors to model real-world behavior and context [37], including inferring information about people and their social networks. Eagle, *et al.* studied and modeled human social structure using mobility data from mobile phones [12,13]. They were able to infer 95% of friendships based on call records and Bluetooth proximity within their participant pool of two interconnected groups at a university. In that study, inferences were made in a very particular setting: the only potential relationships considered were of classmates who were also participants in the study. Our work differs from this project by making inferences about a much larger pool of relationships.

Cranshaw, *et al.* [9] looked at how to use location data alone to infer friendships on Facebook. Using features extracted from the location data, they created a machine learning classifier that achieved 92% accuracy in making a binary classification for someone appearing in a person's list of Facebook friends. Similar to the reality mining work, this work also focused primarily on a campus population, and the analysis could only include relationships where both people in the dyad were participants using the study system. Both of these projects demonstrate the importance, and also the challenge, of using collocation to infer the presence of a social relationship. In our work we focused on trying to be as complete as possible with identifying a user's strong relationships, rather than identifying a subset of the participant's strong relationships. Our error analysis interviews highlight the necessity of collocation for being able to infer some of a person's strong relationships.

Researchers have attempted to use communication data to model social graphs. Past work developed techniques for inferring different groups from email and from online social network usage [5,11,14,22,32]. Past work also looked at inferring strength of ties based on communication patterns [4,16,46]. For example, Xiang et al. [46] developed models that would infer the strength of ties between individuals on LinkedIn. Other work looked at using large quantities of mobile phone call logs to model social structure and persistence of ties over time [20].

Past work has also looked at models to infer the life facet (family, work, or social) of contacts using smartphone logs [29], achieving 90% accuracy for contacts that had at least one communication. Our work uses similar machine learning features. Our core contribution is our investigation into inferring tie strength, our analyses of which techniques did and did not work well, and interviews to better understand the relationship between people's communication actions and tie strength.

3 Methodology

We wanted to assess how well tie strength can be inferred from data stored on people's

smartphones, using their contact lists, call logs, and SMS logs. We wanted to validate the assumption being used in the research community that communication frequency and duration from a few channels can work as an effective proxy for the strength of a relationship. We collected data from January through April 2012 from participants' Android smartphones and asked them to manually categorize and rate their relationships with their contacts to provide ground truth.

3.1 Participants

We recruited 40 participants (13 male and 27 female) living throughout the United States by posting ads in several nationwide online bulletin boards and websites. We had three selection criteria. First, to avoid privacy concerns with minors, participants had to be 18 years or older. Second, to focus on people who could benefit from a more computationally sophisticated representation of relationships, participants had to use Facebook and have at least 50 friends through the service. Third, to ensure a sufficient amount of log data, participants had to have used the same Android phone for at least six months prior to the study. Participant ages ranged from 19 to 50 years (*mean* = 28.0 years, σ = 8.9). Participants were instructed to complete the ground truthing within two weeks, and were compensated $80 USD.

3.2 Procedure

Participants downloaded our Android app, which copied their contact list, call log, and SMS log to a database file. Participants also downloaded their "friends list" from Facebook. Participants then uploaded these files to our server. By default, Android phones limit the call log to the last 500 calls. This resulted in broad differences in how many days the logs represented (range: 21-369; median: 80; mean: 108). Also, most Android phones typically have a default limit of 200 SMS messages per contact.

Next, participants generated a list 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. These people did not need to be in their phone contacts list or on their Facebook friends list. We selected these groups based on past qualitative work [26,40,44].

From the contacts list, Facebook list, and manually created list, we generated a list of 70 contacts for each participant. These were the contacts we wanted to collect ground truth data for. We limited the list to 70 in order to provide enough ground truth data to be useful, but not so much that it would be overly burdensome and cause participants to drop out. To create this list of 70 contacts, we first included all the people each participant manually listed. Next, we selected the top 15 most frequently communicated with individuals for each of phone, SMS, and Facebook. Participants manually 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, with participants manually identifying duplicates. This process continued until we had a list of 70 distinct names for each participant (hereafter referred to as the *70-contact* list).

8

The inclusion criteria for the *70-contact* list are important for interpreting our analysis. First, we wanted to capture all of a participant's close friends, whether or not the participant had logged a phone call or a text message with that person, or even if that person was not in the contact list at all. Second, we wanted to capture the contacts that the participant had communicated with the most. These characteristics allow us to examine the assumptions that communication is a direct proxy for tie strength: we now have ground truth data for all of the high-communication contacts, and we have identified most or all of the participant's close relationships.

When ground truthing, participants provided demographics for each contact, such as sex, age, and duration of the relationship (see Figure 1). Participants also answered four questions about their relationship with each contact, adapted from Marin and Hampton [24]:

1. How close do you feel to this person?
2. How strongly do you agree with the statement "I talk with this person about important matters"?
3. How strongly do you agree with the statement "I would be willing to ask this person for a loan of $100 or more"?
4. How strongly do you agree with the statement "I enjoy interacting with this person socially"?

Participants answered these questions using a discrete 5-point scale [1,4,10,39]. We used a discrete rather than continuous scale to reduce cognitive load and fatigue. Participants were inputting a large amount of data across a large number of contacts, and we were concerned that the freedom of a continuous slider would be an additional burden.

Participants indicated communication frequency using a 7-point scale ranging from *less than yearly* to *daily*. They rated frequency for face-to-face interaction, overall communication through mobile and online channels, mobile phone call, text messaging, and interaction via Facebook.

To protect privacy, we did not collect the content of SMS messages. However, we did collect descriptive information such as email domain name, first six digits of phone numbers, and city/state/zip code, to use as machine learning features.

4 Results FROM OUR DATA COLLECTION

We excluded four participants from our analysis. They each had fewer than two weeks of data and fewer than 100 phone calls. The following results are based on the remaining 36 participants.

In total we gathered logs for 24,370 phone contacts, 16,940 calls, 63,893 text messages (SMS), and 1,853 multimedia messages (MMS). Note: Android phones can be set to automatically sync the phonebook with online contact lists such as Outlook, Gmail, and Facebook, in which case it attempts to merge identical contacts.

As a first step to explore the validity of using information available on a smart phone (contact list, call logs, and SMS logs) to infer tie strength, we examined and compared the answers for the four tie strength questions that our participants answered. The questions were internally consistent and highly reliable ($\alpha = 0.91$). This allows us to combine the answers from the four questions to form a scale. This is a standard practice that increases the reliability of a measure [17]. Based on this combined score we generated a ranked list of each participant's contacts based on relationship strength.

Next we partitioned each participant's contacts into three levels. We based these levels on previous work by Dunbar *et al*, which finds that "rather than a single or a continuous spectrum of group sizes, humans spontaneously form groups of preferred sizes organized in a geometrical series approximating 3–5, 9–15, 30–45, etc." [47]. They found that the top group represents a person's closest relationships (support group), and the second group represents the next closest set of relationships (sympathy group). The larger sized groups of 50 and 150

Figure 1. Sample webpage for ground truth collection. Each participant answered questions about her relationship with selected contacts.
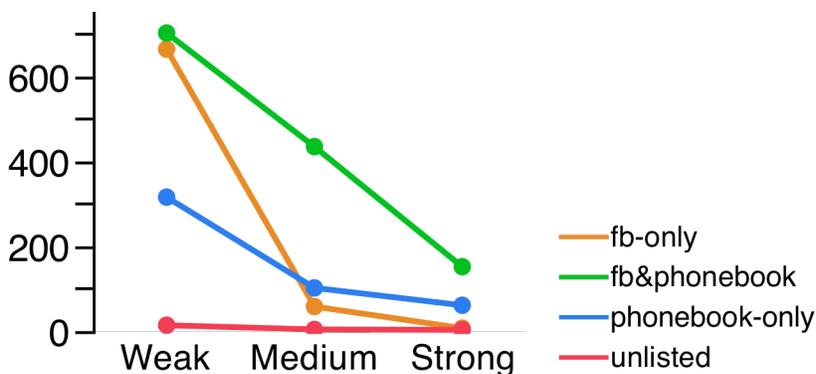
Figure 2. Total number of friends within each tie strength level across all of our participants, separated by the number of people whom the participants had in only in their contact list, only on their Facebook friends list, both in their contact and Facebook lists, or neither.

people are considered to be less stable, and are referred to as clans or regional groupings.

We partitioned each contact list accordingly:

- **strong tie** - the top group (rank 1-5)
- **medium tie** - the middle group (rank 6 – 20)
- **weak tie** - the remaining contacts

Partitioning the data this way also accounts for the individual differences that may have affected a participant's answers to the survey questions.

With these tie strength groupings, we began to investigate communication patterns as a proxy for tie strength. First, we discuss simple features and their relationship to the tie strength groupings. Next, we describe machine learning models for inferring these tie strength levels.

5 Simple Features and Tie Strength

5.1 Contact Source and Tie Strength

The properties of the *70-contact* list allow us to estimate an upper bound for the percentage of a user's close contacts who could be detected from our two contact sources: only Facebook, only the contact list, or both. As Figure 2 shows, overall 99% of people on the *70-contact* list showed up in either a phonebook or Facebook list (range: 95-100%, med: 100%). Overall, 19% of contacts existed only in the phonebook (range: 4-57%, med: 18%); 29% were only in Facebook (range: 0-56%, med: 31%); and 51% were in both (range: 20-90%, med: 52%). Looking across the tie strength categories reveals distinctive trends. We used Spearman's rho (ρ) to measure the non-parametric correlations between tie strength group and presence in the phonebook and Facebook friend list. Being a Facebook-only contact was negatively correlated with tie strength (ρ=-0.32, p < 0.0001). Being a phonebook-only contact was not significantly correlated with tie strength (ρ=0.03, n.s.), although percentage-wise, more of the closer contacts were only in the phonebook. Being a phonebook-and-Facebook contact was positively correlated with tie strength (ρ=0.27, p < 0.0001).

11
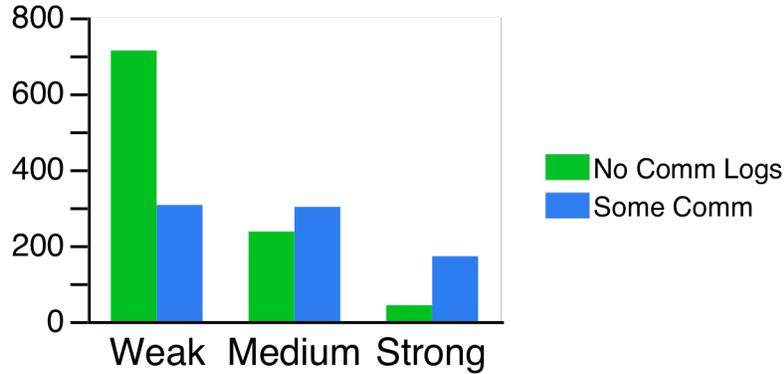
Figure 3. Number of friends in the mobile contact list who exchanged zero or at least one SMS or call with our participants (determined from call log data).

The red points in Figure 2 represent the 21 people that were neither in the phonebook nor Facebook list. These included people whom participants specifically identified as immediate and extended family members; people participants currently live with; work with; feel close to; and do hobbies with. Note that attempts to classify tie strength using call and SMS data alone would completely miss these people.

The orange points in Figure 2 represent Facebook-only contacts and the blue points represent the phonebook-only contacts. 29% of contacts would be missed if using a phonebook-only list to classify tie strength and 19% would be missed if using a Facebook-only list. Both a Facebook-only and a contact-list-only approach would miss some strong ties; however, the Facebook-only approach would miss a notably larger number of strong ties (29% vs. 4%).

5.2 Tie Strength and Phone/SMS Communication

To establish an upper bound for the accuracy of inferring tie strength from phone and SMS communication, we divided the phonebook contacts into two groups by communication history (none vs. some). A reasonable baseline expectation would be that contacts with no communication history would have weak tie strength. Figure 3 shows that most contacts with at least one communication in the dataset tend to have higher levels of tie strength. Additionally, as the tie strength level increases, the percentage of contacts with some communication with the participant also increases ($\rho=0.35$, $p < 0.0001$). Still, several contacts with strong tie strength have no communication history in the dataset. Thus, attempts to classify tie strength using only call and SMS data could not correctly classify these contacts.

Having at least one communication in the call and SMS logs increases the likelihood of a contact having stronger tie strength. However, this is not an absolute rule. There are several counter-examples in both directions - strong tie contacts without communication history and weak tie contacts with it.

Based on the performance of these features, we investigated the relationship between communication frequency and duration with respect to tie strength. Figure 4 plots total call duration and total number of SMS messages against the total count of calls. We expected call duration would be longer with lower frequency for some close contacts. We also expected that call duration might be shorter with higher frequency for other close contacts. For example, a person might call a parent or an old friend infrequently, but chat for a while each time. Conversely, one might call a roommate often to coordinate but not to have lengthy conversations.

As expected, contacts with more frequent or longer duration communications were more often labeled in the higher tie strength levels. Number of calls, duration of calls, and number of SMS are all significantly and positively correlated with tie strength (ρ = 0.42, 0.43, and 0.20, respectively, all p < 0.0001). Surprisingly, many people in all tie strength levels had very
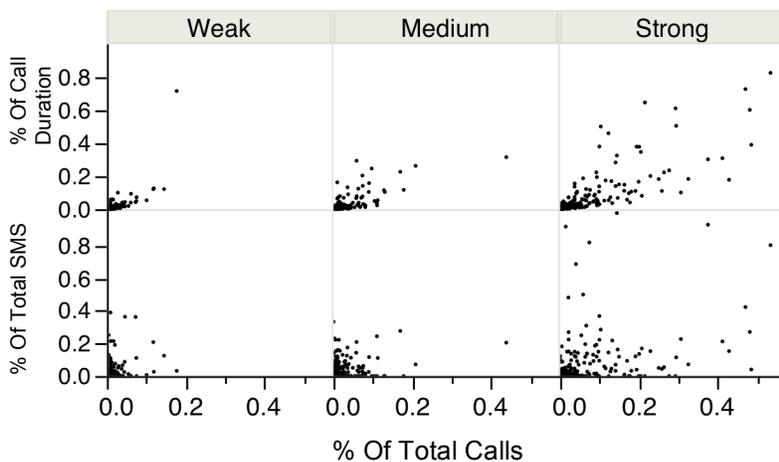
Figure 4. TOP: Percentage of a participant's total talk time spent with each contact vs. percentage of total calls a participant had with each contact. BOTTOM: Percentage of total calls a participant had with each contact vs. percentage of SMS conversations a participant had with each contact. Both are separated across tie strength levels. Each point represents a contact-participant pair who had at least one call or SMS in the dataset. Longer calls, more frequent calls, and more frequent SMS are most common in the strong tie strength categories, but there are also close contacts with low levels for both attributes.

little communication. Lower-tie contacts generally had few calls and short durations. For the higher tie strength levels, the ranges increase for number and duration of calls, but a clump of few-and-short contacts persist.

5.3 Summary

In this section, we established a basic upper bound of accuracy for inferring tie strength with smartphone communication logs. The data shows that using Facebook as the only data source would miss 29% of strong ties, either because they are not Facebook friends, or because these contacts do not use Facebook at all. Next, there are some strong ties without any record of communication within the phone logs. Finally, while communication frequency and duration of calls can help indicate strong tie strength, low frequency and duration are not clear indications of weak tie strength.

These high-level trends are consistent with tie strength theory: more communication on more channels indicates a strong tie. However, our dataset has a number of counterexamples, pointing to critical challenges for automatically inferring tie strength from communication behavior.

6 Classifying Tie Strength

Next we developed a model to classify tie strength based on the call and SMS log data.

6.1 Features Used for Characterizing the Data

We defined a total of 153 machine learning features, based on Min *et al.* [29] (see Table 1): 17 from the contact list, 66 from call logs, 36 from SMS logs, and 34 from combined calls and SMS.

Phonebook contact-profile features (profile features): We expected that similarity between a participant and a contact would indicate strong tie strength. We used two metrics for similarity: same last name (likely to be family) and same first six digits of phone number (same area code + exchange might indicate that phones were purchased at the same time). We also captured the labels on each contact's phone numbers, such as *home*, *work, mobile,* and *other*. This is a predefined set in Android.

Modality		Variables	Description
Survey		is Facebook friend	Available from SNS
Profile features		Similarity of {last name, email domain, company name, phone number}; has {photo, ringtone, webpage, note, affiliation, {work, 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 1. 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.

Very few participants had their own or their contacts' postal-address in the phonebook, so we simply checked if the contact's postal address existed, and used a binary feature for each address label (*home*, *work*, and *other*). We also had additional binary features for additional contact information, such as ringtones, notes, or if the contact was marked as a favorite.

We calculated the completion of each profile as a weak proxy for the effort made to input contact information. Interestingly, we found that the majority of contacts had less than 3 profile attributes, suggesting that more contact information might be a useful signal.

We modeled communication patterns with several categories of features:

- *Intensity and regularity:* The number of and duration of communications has been used to infer tie strength in past work [21,39]. We modeled this factor using features like total number and total duration of calls.
- *Temporal tendency*: In their friends-acquaintances work, Eagle and Pentland observed the temporal tendency in contacting people [12]. For example, calling particular contacts at different times of day and days of the week.

- *Channel selection and avoidance*: People favor a certain communication medium based on the person they are communicating with [27]. We modeled this using features like the ratio between SMS and phone calls.
- *Maintenance cost*: Roberts and Dunbar [39] claim that people apply different amounts of effort in maintaining different kinds of relationships, where effort can be evaluated based on the time to last contact. To model maintenance cost, we measured the number of communications in the past two weeks (short-term view) and in the past three months (relatively long-term view).

Using these factors, we developed a number of features, shown in Table 1 as Call logs, MSG logs (SMS), and COMM (call log + SMS). For messages, we applied a threshold of a 20 minute response time for labeling group messages as a conversation, based on past findings [3].

6.2 Inferring Tie Strength Using Communication Logs

Using all of the features described above, we tested how well a model could infer tie strength. The nature of tie strength poses a challenge for building this model. Tie strength could be treated as a numeric class value based on the answers to the tie strength questions. However, the difference between a rating of 1 and 2 is not necessarily equal to the difference between a rating of 2 and 3. Additionally, our early iterations treating this as a continuous value tended to push scores closer to the middle, with very few people classified as being weak ties. Therefore, we used the tie strength levels of *very strong tie, medium strong tie,* and *weak tie* as nominal class values in these models (explanation of these categories on page 4).

We evaluated our models using the Weka Toolkit's [49] implementation of a support vector machine (SMO). We conducted a leave-one-participant-out cross-validation (each fold contained data from one participant). This prevents any anomalies within a particular participant's data from causing a performance overestimate.

We trained 9 models in total, varying two aspects of input data. First we varied what the model was classifying (first column of Table 2):
- *3-class*: classifies ties as strong, medium-strong, or weak
- *2-strong*: binary classifier that combines medium-strong and weak ties into one class
- *2-medstrong*: binary classifier that combines strong and medium-strong ties into one class

We also varied the input data for the classifier (second column of Table 2):
- *all* includes all contacts on the *70-contact* list
- *contactlist* includes only contacts from the *70-contact* list who appear on the user's mobile phone phonebook (see Figure 2)
- *somecomm* includes only contacts from the *70-contact* list with at least one logged SMS or call (see Figure 3)

Classification results vary considerably (see Table 2), ranging from 46.28% (κ=0.179), to 91.55% (κ=0.361). The Kappa statistic measures the agreement between predicted and observed categorizations, correcting for agreement that occurs by chance.

				Strong ties		Medium strong ties		Weak ties	
Class Cond	Dataset	Accuracy	Kappa	Precision	Recall	Precision	Recall	Precision	Recall
3-class	all	69.60%	0.279	0.503	0.399	0.399	0.209	0.759	0.907
3-class	contactlist	61.11%	0.251	0.491	0.423	0.414	0.242	0.677	0.845
3-class	somecomm	46.28%	0.179	0.449	0.473	0.440	0.425	0.496	0.498
2-strong	all	91.55%	0.361	0.537	0.323			0.937	0.973
2-strong	contactlist	88.64%	0.361	0.545	0.343	⟶		0.914	0.961
2-strong	somecomm	75.46%	0.297	0.480	0.432			0.829	0.855
2-medstrong	all	75.00%	0.367	0.693	0.420			0.764	0.920
2-medstrong	contactlist	68.06%	0.317	0.683	0.460	⟵		0.680	0.843
2-medstrong	somecomm	63.11%	0.192	0.707	0.724			0.488	0.467

Table 2. The results of 9 classifiers constructed using SMO. The prediction classes are tie-strength categories. For 2-strong, the middle and low tie strength classes are combined and for 2-medstrong the middle and high classes are combined.

	Classified As				
	3-class-all			2-strong-all	
Actual	S	M	W	S	W
Strong	89	*57*	*77*	72	*151*
Medium	*61*	125	*413*		
Weak	27	*131*	1540	*62*	2235

Table 3. Confusion matrix for 3-class-all and 2-strong-all models. These classifiers achieved a mean accuracy of 69.6% and 91.6% respectively, with very low recall for classifying strong ties.

Table 2 reveals clear trends. First, within each of the class conditions, classifiers perform best for *all*, second best for *contactlist* and worst for *somecomm*. Figure 2, Figure 3, and Figure 4 provide some insight into these performance results. Most of the contacts who are not in the contact list (and thus are excluded from *contactlist* models) or who have no communication history (and thus are excluded from the *somecomm* models) are not strong ties, and thus are easier to classify. As a result, the models that include them show improved performance.

The most successful class condition is *2-strong*, followed by *2-medstrong*. *3-class* performs the worst. This is typical of multi-class models, which usually take a performance hit compared to binary classifiers. The greatest confusion comes from *medium* ties. Table 3 shows that the majority of medium ties were misclassified as weak.

Most striking, *as many strong ties were incorrectly classified as weak* as were correctly classified as strong. This is perhaps unsurprising, given how many strong ties look like weak ties on measures of call frequency and duration (Figure 4). Importantly, these misclassifications demonstrate the inherent weakness of using call and SMS logs to predict tie strength, and thus the problem with using those logs as direct proxies for tie strength. This result is even more pronounced in recall values for the *strong* tie class of the *2-strong* models in Table 2. The *2-strong-all* model, the model with the highest overall performance, only detects 1/3 of all strong ties.

6.3 Characterizing Model Performance

Depending on the intended use of the classifications, the errors from these models range from

Features with respect to each contact	Info Gain
Total count of calls and SMS	0.198
Total count of calls	0.180
Total duration of all calls	0.163
Total count of days of communication	0.141
% of days with 1+ calls	0.136
% of calls that were on Saturday	0.134
% of call time that was on Saturday	0.120
% of calls that were on Sunday	0.118
% of calls that were before noon	0.115
% of call time that was on Sunday	0.114
STD DEV of outgoing call length	0.112
% of Friday communication that was calls	0.110
% of calls and SMS that were on Saturday	0.109
Max count of incoming calls per week	0.106
% of weekday calls that were long	0.104

Table 2. Top 15 most predictive features using the Information Gain attribute selection with the Ranker search algorithm.

inconsequential to harmful. For example, the classifiers have moderate precision for classifying strong ties, but low recall. They have moderate to high precision for classifying weak ties, with very high recall (except for the *somecomm* models). These models would be useful for applications that want to separate strong ties from weak ties, with some flexibility toward medium ties. However, any application using these models should be robust to misclassifications of strong ties - for all contacts correctly classified as strong there will be at least as many incorrectly classified as weak.

6.4 Best Features for Inferring Tie Strength

The information gain for each feature identifies the most useful features for inferring tie strength. Table 4 shows the top 15 results of an *Information Gain* attribute selection using the ranker search algorithm. This reveals the most valuable features for inferring tie strength.

These results echo earlier findings on the importance of frequency and duration of communication. In the aggregated data, frequency of communication is most important. Other important aspects of the top features include: proportions of calls vs. SMS, the time of day, and day of the week. The proportion of calls initiated also appears to be a factor indicating tie strength.

7 Error Analysis Participant Interviews

Motivated by the low recall of the *strong* tie class in these models, we conducted semi-structured interviews with 7 of our participants. For each participant, we selected 5 to 10 contacts they had labeled as strong ties that we misclassified as weak ties, resulting in 58 contacts being investigated. We asked participants several open-ended questions about the

nature of their relationship and communication with each selected contact. For example:

- When and how did you meet this person?
- What led to this being a close relationship?
- Has anything changed between the time that you became close and now?
- Was there anything different about the channels that you used to communicate with this person or the frequency of communication that you used with this person between then and now?

We coded participants' responses for themes to provide insight into the misclassifications.

7.1 Interview Results

Several themes surfaced that help explain the apparent discrepancy between communication frequency and tie strength. We present them in two categories: *Communication Channel* and *Relationship Evolution.*

Communication Channel

We used to talk on the phone more when we first became close (7 of 58 contacts). In these cases, participants indicated that they spoke on the phone more frequently before, but that they speak on the phone less frequently now, mostly just to catch up. In some cases, this seemed to be a result of a change in life stage (either for the user or for their contact) and/or a change in their geographic location, replicating findings from prior work [41]. For example, one participant complained that he used to keep up with a friend much more regularly before that friend got married, and now they hardly speak at all. Change in life stage and change in geography are discussed more in the *Relationship Evolution* section below.

Other contacts in this category appear to be in relationships in decline, yet the feeling of closeness lingers. One participant spoke about reaching out to a friend multiple times without reciprocity: "I'd like to be friends, but it doesn't work unless we both put in the effort."

In-person communication (11 of 58 contacts). Participants also identified people whom they mostly interacted with in person. A contact's close proximity to the home seems to play an important role. One participant described talking to her neighbor opportunistically, when they see each other. Another detailed how she spoke with her 11-year-old son regularly, just not over the phone. Three participants described friends from classes and their dorm with whom they spoke when they saw each other.

Extended family often fell into this category. Many participants reported primarily speaking with parents, siblings, and other family members in person. In one case, a participant reported going to her parents' house a couple times per month, but mostly not calling her dad on the phone. In these cases, lack of communication logs did not mean lack of effort in maintaining the relationships. In discussing these contacts, some participants specifically mentioned making an effort to travel once a year to see each other, or making a special effort to get together when they do happen to be in the same place.

Other communication channels (25 of 58 contacts). For some strong ties, participants noted

that they communicate regularly, but not via phone calls or SMS. For several participants, communication with a specific contact happened almost exclusively using Facebook. Other participants used instant messenger, email, Skype, or SMS replacements such as WhatsApp to stay in touch with close contacts.

Relationship Evolution

Different location or different life stage (27 of 58 contacts). When asked what was different about their relationship between when they became close and now, many participants responded immediately that either they or their contact had moved. As in the literature [41], participants said that with the change in geography, the communication frequency had changed, but not the perception of closeness. In many cases the move was triggered by a change in life stage (e.g., going to college or grad school, graduating, getting a new job). However, even without moves, a significant life stage change could trigger a communication change on its own (such as getting married or having a first child).

Family is just complicated (17 of 58 contacts). Many misclassified participants were family members. Several participants described specific familial relationships from the perspective of obligation, which hinted at a greater underlying complexity. For example, one participant said that she refused to take her grandmother's phone calls, stating that she calls too frequently and repeats herself. Yet, the participant still reported feeling very close to her grandmother. Another participant, the mother of an 11 year old, said "of course I am close to him," but that it is not necessary for them to talk on the phone. Another participant said her uncle was "definitely close, but he's different from the other close people. He's that really strict uncle that wants to tell me how to live my life, so I don't talk to him too much, maybe every couple months."

7.2 Interview Summary

These interviews highlight the limited effectiveness of the tie strength models we generated. A major issue is the temporality of a relationship. In particular the circumstances under which two people became close are not necessarily the same as the current circumstances of the relationship, even if the two people remain close. The communication logs only capture relatively recent behavior. Therefore, they do not contain the data that indicates a strong long-term relationship. The other main component is that there remains a large amount of interpersonal interaction that happens outside of phone calls and text messages, including communication in other media, as well as face-to-face communication. Call and SMS-based models do not account for this.

8 Discussion

Our work investigates the growing practice of using communication frequency and duration as a proxy for social tie strength. While the social psychology theory identifies that communication frequency and long durations across all communication channels indicate

strong ties, our community has used behavior across a few communication channels and over relatively short time windows as a tie strength proxy. We specifically wanted to know if the call and SMS logs stored on a smartphone held the information necessary to infer tie strength.

8.1 Communication Is an Indicator of Tie Strength, But…

Our results support the tie strength theory literature, showing a strong relationship between tie strength and communication patterns [16,39]. Higher levels of communication frequency, call duration, and, in particular, communication initiated by the phone's owner are all indicators of increased tie strength. However, *the converse is not true.* Low levels of communication do not accurately predict weak ties: our participants had many strong ties who they rarely called or SMSed.

Our interviews probing strong ties with low levels of communication revealed several explanations for this pattern that we believe pose fundamental challenges for automatically inferring tie strength.

First, a person's communication via phone and SMS is not representative of all of their communications. Interactions happen through many other channels (e.g., Skype, instant messenger, landline phones), in some cases replacing communication via phone or SMS. To address this issue, systems will need to collect data from more communication channels. This challenge is multi-dimensional. First, beyond a couple of obvious additional sources (email, Facebook), systems are likely to face diminishing returns when adding additional data sources. For example, some people use Skype, while others use Google Hangouts. Similarly, there are many text message replacement apps (e.g., WhatsApp, GroupMe, Kik). The number of communication channels continues to grow, and there is some speculation that rather than converge, communication channels will become even more diverse over time. Furthermore, people have different preferences for which channels they use and what they use them for. Finally, people switch between services as new ones become trendy, or when friends adopt different services. This highlights another challenge: for many of these services, there is no infrastructure that allows access to this data. For it to be possible to collect data from these additional services, these applications need to support some kind of API.

Second, face-to-face communication remains a primary form of communication for some very close contacts. Capturing this kind of communication is difficult - it is not simply a matter of connecting to an API or setting up a screen scraper. Instead, this will likely require a unique technical solution. The primary challenge is making the solution as lightweight as possible. One approach, from the Reality Mining project is to monitor when two users' Bluetooth radios were within range of each other [12]. This solution unfortunately requires every person to keep their Bluetooth on all the time, to be discoverable to other people, and for each Bluetooth ID to be registered so that it is associated with the right person. Furthermore, co-location does not necessarily mean there is face-to-face communication, as in the case of the familiar stranger [28]. A lightweight, robust solution for inferring face-to-face communication remains an open challenge.

Third, strong ties may form in an earlier life stage and persist across stages even as

communication frequency diminishes. As Spencer and Pahl point out [40], colocation is important in the early stages of a friendship to find commonalities and build the bonds that compose friendship. However, these kinds of interactions are not as essential for maintaining an already strong relationship. For example, close friends in high school may chat on the phone every night. After leaving home and moving on to a new stage of life, they may not communicate as frequently. Yet, each may still feel quite close to their "old" friend. The loyalty of the strong tie may far outlast the underlying behavior that developed the tie in the first place. With short communication and colocation logs, it is impossible to tell if a contact has a small amount of communication due to lack of closeness or if the pair is maintaining a relationship that is already close.

It is very likely that log data from any computer-mediated communication will have similar challenges in predicting tie strength. It seems the only way to automatically detect these types of already-close relationships is through communication logs that extend back to when those close relationships first began.

A child born today could, potentially, have a full lifespan of computer-mediated communication logs. This would necessitate several changes in infrastructure, including storage of longer communication history (the devices in this study limited storage to the last 500 calls and 200 SMS messages), standardization of APIs to access the data, and the creation of a unified structure for storing user data and maintaining history as users change devices and services.

For a relationship that is already a strong tie, perhaps any communication, no matter how little, is an indicator of the continued strength of that relationship. Differentiating these maintenance communications from other communications would be a meaningful step forward. Time from last communication is a good start for modeling maintenance costs [39], but we may need different metrics for understanding what communication should be interpreted as maintenance in the first place.

Access to long-term data will raise other questions: what is the half-life of a close relationship? Can close relationships lie dormant indefinitely and be revived at any time? Also, when does a close relationship reach a level where it does not require the same level of communication frequency, and can this be inferred automatically? This work lays a foundation for exploring these questions in further detail.

8.2 Limitations for Call and SMS as a Proxy for Tie Strength

A breadth of recent research has assumed that call and SMS behavior is a good proxy for tie strength [8,31,36,43]. Our results indicate that this assumption can be problematic. These results are an important point of consideration for any researchers who wish to draw claims using call and SMS data as a tie strength proxy. While theoretical literature supports the relationship between communication frequency and duration and tie strength [21], these communications should not be operationalized only through call and SMS logs. In particular, this operationalization will miss a significant portion of close relationships.

8.3 Eliciting Contacts to Label

In this work we purposefully asked our participants to list all of their close contacts, regardless of whether or not they shared any calls or SMS messages. Additionally, we explicitly asked if the most frequently called and SMSed friends were close or not. We realized that gathering data this way would make the classification task harder: we expected that close people with no calls or SMS messages would be misclassified as not close. Similarly, by explicitly including contacts that have had more frequent communication regardless of tie strength, we created an even greater challenge for our model. Instead of selecting contacts at random, our method specifically targeted strong ties (that were likely to have extremely different communication patterns) and contacts with frequent communication (that were likely to vary in tie strength). Though this data collection method reduced classification accuracy, it better represented the range of communication patterns among strong ties. By doing so, this dataset highlights the limitations of using only communication data to infer tie strength. With this study design, we can quantify how many close contacts are likely to be completely missed, and how many not-close contacts would be difficult to distinguish from close ones.

This is an important difference that distinguishes this work from the significant contribution by Gilbert and Karahalios, who sampled Facebook contacts only, and did so at random [16]. Our data collection method allows us to say with much higher confidence that we are classifying all of a participant's close contacts, instead of just their close contacts with whom they communicate using their cell phone. Despite these differences, the accuracy achieved in this work for binary classification (91.6%) was similar to the accuracy they were able to achieve for a binary prediction (85%). Our study design also enabled us to quantify the close contacts that are not captured by calls and SMS messages, and also to conduct follow-up interviews to further investigate the model misclassifications.

9   Conclusion

In this paper, we evaluated the ability of call and SMS logs stored on users' smartphones to infer tie strength. We achieved 91.6% accuracy for binary classification. We also found that, contrary to the assumptions made in past work, using phone communication frequency and duration as a proxy for tie strength is not fully effective for distinguishing between a user's close and less close relationships. Interviews with our participants revealed several explanations for low frequency, short duration communication with close contacts – these explanations indicate fundamental limitations when using communication logs to infer close relationships.

10 Acknowledgments

References

1.   Ackerman, J.M., Kenrick, D.T., and Schaller, M. Is friendship akin to kinship? *Evolution and Human Behavior 28*, 5 (2007), 365–374.

2.   Allen, T. Managing the flow of scientific and technological information. 1966.

3.   Battestini, A., Setlur, V., and Sohn, T. A large scale study of text-messaging use. *MobileHCI*, (2010), 229–238.

4.   Burke, M. Reading, Writing, Relationships: The Impact of Social Network Sites on Relationships and Well-Being. 2011.

5.   De Choudhury, M., Mason, W.A., Hofman, J.M., and Watts, D.J. Inferring relevant social networks from interpersonal communication. *WWW*, ACM Press (2010), 301.

6.   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, 2009.

7.   Conti, M., Passarella, A., and Pezzoni, F. A model for the generation of social network graphs. *2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, IEEE (2011), 1–6.

8.   Conti, M., Passarella, A., and Pezzoni, F. A model for the generation of social network graphs. *WWMMN*, IEEE (2011), 1–6.

9.   Cranshaw, J., Toch, E., Hong, J.I., Kittur, A., and Sadeh, N. Bridging the Gap Between Physical Location and Online Social Networks. *Ubicomp 2010*, (2010).

10.   Cummings, J.N., Lee, J.B., and Kraut, R. Communication technology and friends during the transition from high school to college. In R.E. Kraut, M. Brynin and S. Kiesler, eds., *Computers, phones, and the Internet: Domesticating information technology*. Oxford University Press, New York, 2006, 265–278.

11.   Danezis, G. Inferring Privacy Policies for Social Networking Services. *AISec*, (2009).

12.   Eagle, N., Pentland, A., and Lazer, D. Inferring Social Network Structure using Mobile Phone Data. *The National Academy of Sciences (PNAS) 106*, 36 (2009), 15274–15278.

13.   Eagle, N. and Pentland, A. Social serendipity: Mobilizing social software. *IEEE Pervasive Computing 4*, 2 (2005), 28–34.

14.   Fang, L. and LeFevre, K. Privacy wizards for social networking sites. *WWW*, (2010), 351–360.

15.   Festinger, L. Informal social communication. *PsychRev 57*, 5 (1950), 271–282.

16.   Gilbert, E. and Karahalios, K. Predicting tie strength with social media. (2009), 211–220.

17.   Gliem, J.A. and Gliem, R.R. Calculating, Interpreting, And Reporting Cronbach's Alpha Reliability Coefficient For Likert-Type Scales. 2003. https://scholarworks.iupui.edu/handle/1805/344.

18.   Granovetter, M. The strength of weak ties. *American journal of sociology*, 1973. http://www.journals.uchicago.edu/doi/abs/10.1086/225469.

19.   Granovetter, M. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory 1*, (1983), 201–233.

20.   Hidalgo, C.A. and Rodriguez-Sickert, C. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications 387*, 2008, 3017–3024.

21.   Hill, R.A. and Dunbar, R.I.M. Social network size in humans. *Human Nature 14*, (2003), 53–72.

22.   Jones, S. and O'Neill, E. Feasibility of Structural Network Clustering for Group-Based Privacy Control in Social Networks. *SOUPS*, (2010).

23.   Kim, J. and Zimmerman, J. Cherish. *CHI '06 extended abstracts on Human factors in computing systems - CHI EA '06*, ACM Press (2006), 953.

24.   Marin, A. and Hampton, K.N. Simplifying the Personal Network Name Generator. *Field Methods 19*, 2007, 163.

25.   Marsden, P. V and Campbell, K.E. Measuring tie strength. *Social Forces 63*, 1984, 482–501.

26.   McCarty, C. Structure in personal networks. *Journal of Social Structure 3*, 1 (2005).

27.   Mesch, G.S. Social context and communication channels choice among adolescents. Computers. *Human Behavior 25*, (2009), 244–251.

28.   Milgram, S., Sabini, J., and Silver, M. *The individual in a social world: Essays and experiments (2nd ed.). McGraw-Hill series in social psychology.* .

29.   Min, J.-K., Wiese, J., Hong, J.I., and Zimmerman, J. Mining smartphone data to classify life-facets of social relationships. *CSCW '13*, ACM Press (2013), 285.

30.   Miritello, G., Moro, E., Lara, R., et al. Time as a limited resource: Communication strategy in mobile phone networks. *Social Networks 35*, 1 (2013), 89–95.

31.   Miritello, G., Moro, E., Lara, R., et al. Time as a limited resource: Communication strategy in mobile phone networks. *Social Networks 35*, 1 (2013), 89–95.

32.   Neustaedter, C., Brush, A.J.B., Smith, M.A., and Fisher, D. The Social Network and Relationship Finder: Social Sorting for Email Triage. *Conference on Email and Spam (CEAS 2005)*, (2005).

33.   Newcomb, T. *The acquaintance process*. New York: Holt, Rinehart, & Winston, 1961.

34.   Odom, W., Zimmerman, J., and Forlizzi, J. Teenagers and Their Virtual Possessions: Design Opportunities and Issues. *In Proc. CHI '11*, .

35.   Onnela, J.-P., Saramäki, J., Hyvönen, J., et al. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America 104*, 18 (2007), 7332–6.

36.   Onnela, J.-P., Saramäki, J., Hyvönen, J., et al. Structure and tie strengths in mobile communication networks. *PNAS 104*, 18 (2007), 7332–6.

37.   Pentland, A. *Honest Signals: How They Shape Our World (Bradford Books)*. The MIT Press, 2008.

38.   Roberts, S.G.B., Dunbar, R.I.M., Pollet, T. V., and Kuppens, and T. Exploring variation in active network size: Constraints and ego characteristics. *Social Networks 31*, 2 (2009), 138–146.

39.   Roberts, S.G.B. and Dunbar, R.I.M. Communication in social networks: Effects of kinship, network size, and emotional closeness. *Personal Relationships 18*, 3 (2011), 439–452.

40.   Spencer, L. and Pahl, R. Rethinking Friendship: Hidden Solidarities Today. *Princeton University Press*, 2006.

41.   Spencer, L. and Pahl, R.E. *Rethinking friendship: hidden solidarities today*. Princeton University Press, 2006.

42.   Wang, D., Pedreschi, D., Song, C., Giannotti, F., and Barabasi, A.-L. Human mobility, social ties, and link prediction. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11*, ACM Press (2011), 1100.

43.   Wang, D., Pedreschi, D., Song, C., Giannotti, F., and Barabasi, A.-L. Human mobility, social ties, and link prediction. *KDD '11*, ACM Press (2011), 1100.

44.   Wiese, J., Kelley, P., 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. *Ubicomp 2011*, (2011).

45.   Wiese, J., Kelley, P.G., Cranor, L.F., Dabbish, L., Hong, J.I., and Zimmerman, J. Are you close with me? are you nearby? *In Proc. UbiComp '11*.

46.   Xiang, R., Neville, J., and Rogati, M. Modeling relationship strength in online social networks. *WWW*, (2010), 981–990.

47.   Zhou, W.X., Sornette, D., Hill, R.A., and Dunbar, R.I.M. Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B: Biological Sciences 272*, 2005, 439.

48.   Zipf, G. *Human behavior and the principle of least effort*. Addison-Wesley, Cambridge, MA, 1949.

49.   Weka 3: Data Mining Software in Java. .