"You Never Call, You Never Write": Call and SMS Logs Do Not Always Indicate Tie Strength

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

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. 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 36 participants. Consistent with theory, we found that frequent or long-duration communication likely indicates a strong tie. However, the use of call and SMS logs produced many errors in separating strong and weak ties, suggesting this approach is incomplete. Follow-up interviews indicate fundamental challenges for inferring tie strength from communication logs.

Author Keywords
Tie strength, smartphone, social graph

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

INTRODUCTION
Logged communication data collected by smartphones offer a potentially rich way for computer systems to gain more 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 [20]; online identity authentication [21]; virtual possession collections that automatically surface based on the social context [30]; and even services that automate the management of privacy settings based on the influence of relationship in people's willingness to share [39].

A wealth of work in social psychology has found that 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 [16,23]. Subsequent research has operationalized this theory by using call frequency and duration as a proxy for tie strength [7,27,31,37].

The theory relating communication frequency to tie strength is based on communication across long periods of time and across all possible channels, including face-to-face. Many recent and highly cited papers, however, have operationalized tie strength using data from one or two communication channels and using small time windows when compared to a person's lifetime [7,27,31,37]. The idea that you can model tie strength with relatively sparse data is tantalizing but dangerous. If it works, it is a discount approach that can easily be adopted by many applications where call logs are readily accessible. However, linking together communications from more data sources is challenging and error-prone. Furthermore, if sparser communication data does not accurately infer tie strength, then our community's ongoing adoption of this untested assumption may lead to many kinds of errors.

To investigate if tie strength can be inferred from sparser communication data sets, we conducted a study based on the call and SMS logs stored on smartphones that had been in use for at least six months. We gathered data from 36 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 seventy of their contacts and friends.

The data confirms that high-frequency or long-duration communication is a useful signal for identifying strong ties. However, communication frequency and duration are a noisy signal, producing many errors. In an attempt to reach beyond the simple proxies of frequency and duration, we employed a set of 153 features and developed 9 machine learning models to identify any complex signals in the call logs that might indicate tie strength. 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 is a useful signal for identifying strong ties. However, this high accuracy can be misleading, because the number of strong ties is much lower than the number of weak ties. Despite the high accuracy of the model, over half of strong ties are incorrectly labeled as weak ties, and only half of the classified strong ties are actually strong ties.

CSCW '15, March 14 – 18, 2015, Vancouver, BC, Canada.

http://dx.doi.org/10.1145/2675133.2675143

To understand the cause of these errors, we interviewed a subset 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) face-to-face communication is important, but is not easily observed; and 3) people feel a lingering sense of closeness to friends from a previous stage in their life, though communication has decreased.

In hindsight, these results may seem obvious; interestingly, they are not well represented in the literature. Furthermore, several highly cited papers operationalize tie strength only using call frequency and duration, without acknowledging these limitations, leading to an incomplete construct that may affect their results. While in many cases a more accurate approximation of tie strength may not be feasible, work using call logs to operationalize tie strength should consider these systematic biases and directly address how these limitations affect their claims.

RELATED WORK

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 closeness and kinship influence the size of social networks and communication patterns [34,35]. Some studies have used proximity as a proxy for quantity of social interaction between pairs [13,29], showing that communication frequency drops exponentially as members of a pair move farther apart [2,42].

Granovetter identified four dimensions of tie strength: duration, intimacy, intensity, and reciprocity [17]. 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 [14]. 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 they are in the participant's contact list; and 2) Selecting contacts with whom the participant has frequent communication. By employing this approach, we can 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.

Backstrom and Kleinberg used Facebook data to analyze network properties of users to identify a specific kind of strong tie: a romantic partner [3]. They identified a new measure of tie strength that they call 'dispersion;' the lack of connectedness between two people's mutual friends. Employing machine learning techniques focused on structural and interaction measures, they correctly identified an individual's romantic partner 70% of the time. The call and SMS log data we used lacked these structural features. The logs contain egocentric network information for each participant, not for each of the participant's contacts. However, we take a similar process of examining several specific features individually before combining them in a machine learning model to infer strong ties.

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 [32], including inferring information about people and their social networks. Eagle, *et al.* studied and modeled human social structure using mobility data from mobile phones [11,12]. 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.* [8] 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 used communication data to model social graphs. Past work developed techniques for inferring different groups from email or online social network usage [5,10,28]. Other work looked at inferring tie strength based on communication patterns [4,14,40]. For example, Xiang et al. [40] 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 [18].

Other work has created models to infer the life facet (family, work, or social) of contacts using smartphone logs [26], 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 understand the errors made by the models.

METHOD

We wanted to assess how well tie strength can be inferred from data found on nearly every smartphone: contacts, call logs, and SMS logs. We chose these data sources because we wanted to validate the assumption being used in the research community that communication frequency and duration from these channels can work as an effective proxy for the strength of a relationship. We collected data from participants' Android smartphones and asked them to manually categorize and rate their relationships with individual contacts as our ground truth.

Participants

We recruited 40 participants (13 male and 27 female) living throughout the United States by posting ads in several places: on Craigslist in 6 major US cities, on a nationwide site for recruiting study participants, on a website for posting social relationship research studies, and on a participant pool within our university. We had three selection criteria. First, to avoid privacy concerns with minors, participants had to be at least 18. 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. 55% of our participants were students (graduate or undergraduate), 35% were employed in a variety of professions, and 10% were unemployed. Participant ages ranged from 19 to 50 years (*mean* = 28.0 years, $\sigma = 8.9$). Participants were instructed to complete the ground truthing within two weeks, and were compensated $80 USD. Of the 40 participants, we excluded four participants from our analysis: each had fewer than two weeks of data and fewer than 100 phone calls. Findings are based on the remaining 36 participants.

Procedure

Participants downloaded our Android app, which copied their contact list, call log, and SMS log to a database file. Participants then uploaded this file, in addition to their Facebook friends list, to our server through a custom website that was designed for this study. The entire study was conducted through this website. Participants could stop and resume whenever they wanted, and were given two weeks to complete the entire process. By default, Android phones limit the call log to the last 500 calls and typically have a default limit of 200 SMS messages per contact. This resulted in broad differences in how many days the logs represented (range: 21-369; median: 80; mean: 108).

Participants' contact and Facebook lists were much too long for participants to completely ground truth. Through pilot testing, we found 70 contacts to be a reasonable number for participants to rate before becoming overly burdensome: we wanted to maximize participant retention.

The vast majority of any individual's contacts will be weak ties. However, in this work, we wanted to maximize the likelihood of collecting information on strong ties. To accomplish this, we asked participants to generate a list of contacts that fit specific social categories, regardless of their appearance in the phone contact list or Facebook list. Participants listed five people in each of the following categories: *immediate family*, *extended family*, *people they live with*, *coworkers*, *people they feel close to*, and *people they do hobbies with*. We selected these groups based on past qualitative work that suggests these categories will contain an individuals' strong ties [24,36,39]. This resulted in approximately 25 unique names per participant (some names were repeated across the categories). In addition, we selected the 15 contacts with the highest communication frequency for calls, SMS, and Facebook. 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 many of the participant's strong ties by asking for them directly. If call and SMS communication is a perfect proxy for tie strength, these two groups should be the same.

To fill out the list of 70 contacts, we combined category list and the frequency list, removing all duplicates. To get the list of 70 contacts, we added randomly selected individuals from their phone's contact list and Facebook friend list, with participants manually identifying duplicates due to the challenges of automatic merging using names [38]. This process continued until we had a list of 70 distinct names for each participant (hereafter called the *70-person* list).

Participants provided demographics for each contact in the 70-person list, such as sex, age, and relationship duration. Participants also answered four questions about their relationship with each contact, adapted from [22]:

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"?

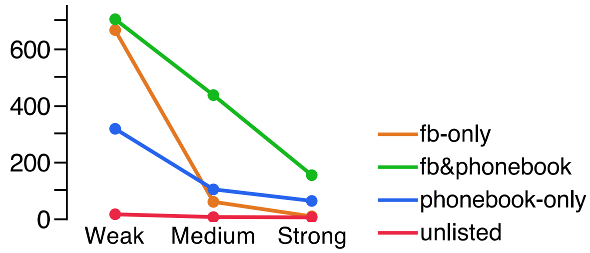

Figure 1. Total number of friends within each tie strength level across all participants, separated by the number of contacts who only appeared in the contact list, only in the Facebook friends list, appeared in both, or neither.

4. How strongly do you agree with the statement "I enjoy interacting with this person socially"?

Participants answered questions using a discrete 5-point scale, following previous work on tie strength [1,4,9,35]. We used a discrete rather than continuous scale to reduce cognitive load and fatigue. Participants provided a large amount of data for many contacts, and we were concerned that the freedom of a continuous slider would be an additional burden. 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.

DATA ANALYSIS

We gathered logs for 24,370 phone contacts, 16,940 calls, 63,893 SMS messages, and 1,853 MMS messages. Note: Android phones can be set to automatically sync the phonebook with online contact lists (e.g. Gmail and Facebook). 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 analyzed the answers for the four tie strength questions our participants answered. The questions were highly reliable ($\alpha = 0.91$). This allows us to add together the answers from the four questions to form a scale. This is a standard practice that increases the reliability of a measure [15]. 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 explored several approaches for identifying these levels. An assessment of the distribution of Z-scores from the combined tie strength metric both across all participants and per-participant revealed no obvious gaps in ratings on which we could split strong and weak ties. Instead, we based these levels on previous work by Zhou *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." [41]. 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 people are considered to be less stable, and are referred to as clans or regional groupings.

In constructing each participant's 70-person list, we took multiple steps to increase the likelihood of capturing many of a participant's closest contacts. Because of this, we assigned the contacts into their respective groups based on the numbers from Zhou *et al*. By handling the data in this way, we are able to normalize out individual differences between participants (e.g. a tendency for some participants to use 3 as the baseline and others to use 1, or a participant's negative reaction to a particular question).

We partitioned each contact list accordingly:

- ***strong tie*** - the top group (rank 1-4)
- ***medium tie*** - the middle group (rank 5 – 19)
- ***weak tie*** - the remaining contacts

In cases where multiple contacts tied for a rank, all of those contacts were assigned to the same tie strength level, resulting in a slight variation in group sizes per participant.

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.

SIMPLE FEATURES AND TIE STRENGTH

Contact Source and Tie Strength

The properties of the *70-person* 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 1 shows, overall 99% of people on the *70-person* 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, $p < 0.001$). Being a phonebook-only contact was not 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, $p < 0.001$).

The red points in Figure 1 represent the 21 people that were neither in the phonebook nor Facebook list, including people whom participants identified as immediate and extended family members; people participants currently live with; work with; feel close to; and do hobbies with. The orange points in Figure 1 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;

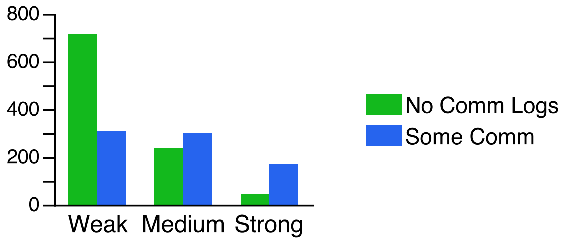

Figure 2. Number of friends in the mobile contact list who exchanged zero or at least one SMS or call with our participants (determined from call log data).

however, the Facebook-only approach would miss a notably larger number of strong ties (29% vs. 4%).

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 2 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 (ρ=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 higher tie strength. However, this is not an absolute rule: there are counter-examples in both directions - strong ties without communication history and weak tie contacts with it.

Next we explored the relationship between communication frequency and duration with respect to tie strength. Figure 3 shows six plots in a grid. The top row shows aggregate call duration (y-axis) against the total number of calls (x-axis) for each contact. The bottom row shows the total number of SMS messages (y-axis) against the total number of calls (x-axis) for each contact. Each column indicates the contact's ground truth tie strength level. Both aggregate duration and frequency are represented as a percentage relative to the total call duration or number of calls per participant. We expected some close contacts to stand out with long call durations, and others to stand out with high frequency. For example, a person might call an old friend infrequently, but chat for a while each time. Conversely, one might regularly make short calls to a roommate to coordinate.

As expected, contacts with more frequent or longer duration communications were more often in the higher tie strength levels. Number of calls, duration of calls, and number of SMS are all positively correlated with tie strength (ρ = 0.42, 0.43, and 0.20, all $p < 0.0001$). Surprisingly, many people in all tie strength levels had very little communication. Weak ties generally had few calls and short durations. For

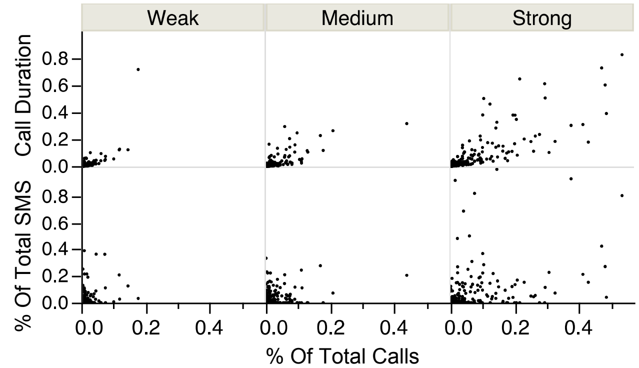

Figure 3. A grid of six plots showing communication frequency and total talk time. The top 3 graphs plot each contact's aggregate call duration (y-axis) against number of calls (x-axis). The bottom 3 graphs plot each contact's number of SMS messages (y-axis) against number of calls (x-axis). For both top and bottom, the columns separate the contacts by tie strength group. The graphs include data for contacts with at least one call or SMS. All numbers are represented as the percentage of a participant's total communication frequency/duration.

strong ties, the ranges increase for number and duration of calls, but a clump of few-and-short contacts persist.

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 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.

CLASSIFYING TIE STRENGTH

While the above findings already indicate significant issues for using call and SMS logs to indicate tie strength, we wanted to explore the possibility that a combination of more subtle features than frequency and duration might indicate tie strength. To explore this prospect, we developed several machine learning models to classify tie strength based on call and SMS log data.

Features Used for Characterizing the Data

We defined a total of 153 machine learning features: 17 from the contact list, 66 from call logs, 36 from SMS logs, and 34 from combined calls and SMS. These features are based on Min *et al.* [26], and more details on the specific features can be found in that paper. These features include:

- *Intensity and regularity:* The number of and duration of communications has been used to infer tie strength in past

				Strong ties		Medium strong ties		Weak ties	
Class Condition	**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-verystrong	all	91.55%	0.361	0.537	0.323			0.937	0.973
2-verystrong	contactlist	88.64%	0.361	0.545	0.343			0.914	0.961
2-verystrong	somecomm	75.46%	0.297	0.480	0.432			0.829	0.855
2-mediumstrong	all	75.00%	0.367	0.693	0.420			0.764	0.920
2-mediumstrong	contactlist	68.06%	0.317	0.683	0.460			0.680	0.843
2-mediumstrong	somecomm	63.11%	0.192	0.707	0.724			0.488	0.467

Table 1. The results of 9 classifiers constructed using SMO. The prediction classes are tie-strength categories. For 2-verystrong, the middle and low tie strength classes are combined and for 2-mediumstrong the middle and high classes are combined.

work [19,35]. 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 [11]. 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 [25]. We modeled this using features like the ratio between SMS and phone calls.
- *Maintenance cost*: Roberts and Dunbar [35] find that people apply different amounts of effort in maintaining different kinds of relationships, based on the time to last contact. To model maintenance cost, we used the number of communications in the past two weeks (short-term view) and in the past three months (longer-term view).

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 [43] 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, varying two aspects of input data. First we varied what the model was classifying (First column of Table 1):

- ***3-class***: classifies as strong, medium-strong, or weak
- ***2-verystrong***: binary classifier that combines medium-strong and weak ties into one class
- ***2-mediumstrong***: 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 1):

- ***all*** includes all contacts on the *70-person* list
- ***contactlist*** includes only contacts from the *70-person* list who appear in the user's phonebook (see Figure 2)
- ***somecomm*** includes only contacts from the *70-person* list with at least one logged SMS or call (see Figure 3)

Classification results vary considerably (Table 1), 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. Table 1 reveals clear trends. First, within each of the class conditions, classifiers perform best for *all*, second best for *contactlist* and worst for *somecomm*. Figures 2, 3, and 4 provide some insight into these results. Most of the contacts who are not in the contact list (thus excluded from *contactlist* models) or who have no communication history (thus excluded from the *somecomm* models) are not strong ties, and thus are easier to classify. As a result, the models that include them perform better.

The most successful class condition is *2-verystrong*, followed by *2-mediumstrong*. *3-class* performs the worst. This is typical of multi-class models, which usually take a performance hit compared to binary classifiers.

The models incorrectly classified more strong ties as weak than were correctly classified as strong (in Table 1, the recall values for the strong tie class are the percentage of strong ties correctly classified). Also, about half of ties that were classified as strong were actually not strong (in Table 1, the precision values for strong ties is the percentage of contacts that were classified as strong ties who were actually strong ties). The plots from Figure 3 offer insight into these errors. These misclassifications emphasize the weakness of using call and SMS logs to infer 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-verystrong* models in Table 1. The *2-verystrong-all* model, the model with the best performance, only detects 1/3 of strong ties.

ERROR ANALYSIS PARTICIPANT INTERVIEWS

Motivated by the particularly 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, (58 contacts total). An error analysis of the data led us to focus on strong ties that were classified as weak. In the error analysis, we referenced tie strength theory to consider communication expectations for medium and weak ties. It is not that we expect people to *only* communicate with strong ties, so the presence of some communication with weak ties is reasonable. However, if participants had more communication with more of their strong ties, the model would have been better able to distinguish between strong and weak ties. This led us to focus on strong ties with little or no communication, rather than weak ties with some communication.

Interviews took place over the phone, lasted about a half an hour, and were recorded to facilitate note taking. We asked participants open-ended questions about the nature of their relationship and communication with each selected contact:

- 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?

Interview Results

We iteratively coded participants' responses about each contact for themes to provide insight into the misclassifications. Several themes surfaced that help explain the 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 [36]. 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 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 [36], participants said that with the change in geography, the communication frequency had changed, but not the perception of closeness. The move was often triggered by a change in life stage (e.g., going to college, graduating, getting a new job). However, even without moves, a significant life stage change could trigger a communication change on its own (e.g. getting married or having a child).

Family is close regardless of communication (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."

Interview Summary

These interviews highlight the limited effectiveness of the tie strength models. 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.

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 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 wanted to know if the call and SMS logs stored on a smartphone held enough information to infer tie strength.

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 [14,35]. Higher levels of communication frequency, call duration, and, in particular, communication initiated by the phone's owner are all indicators of a strong tie. However, we found that when operationalizing this theory with call and SMS logs, the signal is very noisy. Low levels of communication do not accurately identify weak ties: our participants had many strong ties who they rarely called or SMSed. Our interviews probing strong ties with little communication revealed several explanations for this pattern that we believe pose fundamental challenges for inferring tie strength.

First, a person's communication via phone and SMS does not capture 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. Second, face-to-face communication remains a primary form of communication for some very close contacts, but capturing this kind of communication today is difficult. Third, strong ties may form in an earlier life stage and persist across stages even as communication frequency diminishes. Even if we could capture data across multiple channels and do so for longer periods of time, it is not clear that this would be sufficient to improve the models of tie strength.

A breadth of recent and highly-cited research has assumed that call and SMS behavior is a good proxy for tie strength [7,27,31,37]. These contributions do not attempt to identify all strong ties exhaustively. Rather, they only identify strong ties who use a specific communication channel. Our *contactlist* and *somecomm* datasets best match this task. The models for these datasets produce similar errors; they also indicate that communication frequency and duration are an incomplete signal for determining tie strength. While theory supports the relationship between communication frequency and duration and tie strength [19], these communications should not be operationalized only through the call and SMS logs stored on a person's phone.

Alternatives for Identifying Tie Strength

Researchers looking for a way to separate strong ties and weak ties need to consider alternatives to using short term communication logs from one or two channels, such as those available of today's smartphones.

One alternative is to collect data from more communication channels. This approach has several challenges. First, beyond a couple of obvious additional sources (i.e. email, Facebook), researchers 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, people have different preferences for which channels they use and for what purposes, and people switch between services based on fads, or on what services friends are using. Second, many of these services offer no API for accessing this data. Third, correctly linking contact identities across multiple communication sources is non-trivial and error-prone.

Another way of augmenting this process while still using communication data to separate strong and weak ties is to use *a lot more data*; data that extends back to when close relationships first began, which could be on the order of years or even decades. Since this data does not currently exist, the only way to evaluate this is to start collecting the data now and see if it helps several years from now. Current data collection and retention practices are not helping to solve this challenge for researchers. For example, Android devices by default only store the last 500 calls and 200 SMS messages. Furthermore, there are no standard APIs to access one's data, and no unified structures for storing user data and maintaining history as users change devices and services. If this kind of work is ever going to be possible, these practices will have to change.

Investigating message content might also help to improve the separation of strong and weak ties. It is possible that in cases where there is some communication, the content of the communication with strong ties is different from weak ties in a systematic way. A drawback to this approach, and the reason that we did not explore this avenue, is that many people are uncomfortable with the privacy implications of granting content level access to calls and SMS.

Another approach is to try and differentiate relationship maintenance communications with strong ties (which can be less frequent but very important to maintaining the strong tie) from other communication. There are many possible opportunities. One example might be to see whom a person calls or visits when traveling (factoring in time of day to differentiate between a likely work contact versus a social contact). Another example might be to use age or the inferred life stage of individuals and incorporate that into tie strength models. For instance, college students, 40-year-old parents, and senior citizens likely have different kinds of people in their strong ties. This idea would require much deeper investigation into how people's friendships change over time and how life stage affects these relationships.

The most reliable (and the most obvious) option for distinguishing strong and weak ties is to include users in the process through interviews [36], or a survey as we did. Some research has also looked at computer supported tools for collecting this kind of data [33]. The primary challenge here is that, even in the case that labeling is efficient, this approach still requires the time and effort of the user.

The primary drawback to all of these approaches is that in general, researchers who use communication frequency as a tie strength proxy do so because it is easily available. Many of the datasets that are being analyzed were collected and anonymized for a different purpose, often by a third party such as a telecommunications company. These researchers do not have the possibility of collecting more data, or have any access at all to the actual participant. Furthermore, many of these datasets contain the data of far too many users for a non-automated approach to be possible.

Using Communication Frequency as Tie Strength

We expect researchers to continue to use communication frequency as a tie strength proxy because it is now available due to the increasing use of smartphones. Here, we offer some implications for those that make this choice.

Researchers should carefully consider how the imperfect proxy of communication frequency as tie strength limits the strength of their claims. Does the fact that a strong tie might have some in-channel communication (which means that they would be included in an experiment), but has less communication than some weak ties hurt the strength of a claim being made on that data? It will depend on the claims being made, and to what extent those claims rely on a clear separation between strong and weak ties.

One solution for researchers in these situations is to modify the claims in their papers so that instead of relating their claims to *tie strength*, they relate the claims directly to *communication frequency*. For example, the existing work [7,27,31,37] that equates tie strength and communication frequency are valuable contributions. The issue with these works is that explaining their findings in the context of "tie strength", while convenient, gives the false impression that the work is based on a reliable measure of tie strength. This can negatively impact the reader's ability to correctly interpret their findings. If tie strength is important to an argument, researchers should also explain how they believe tie strength and communication frequency are related to each other within their dataset, and should explicitly identify that communication frequency is a limited proxy.

In this initial work, we have not yet explored the possibility of systematic per-user differences based on demographics, behavioral characteristics, or life stage, that affect our ability to separate strong ties from weak ties. If any such effects exist, they may well have an impact on the claims that can be drawn from using communication frequency to classify tie strength. Similarly, it is conceivable that there are other dimensions of interpersonal relationships that communication frequency is capable of detecting. Perhaps through this process, we can further our definition and understanding of the nuances of tie strength as a concept.

CONCLUSION

Having a computational model of tie strength could be useful for a number of domains. Past work has used call log data as a proxy for tie strength. However, our analysis of 36 participants' data suggests that this operationalization of tie strength is incomplete, missing more strong ties than it correctly identifies. 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.

ACKNOWLEGEMENTS

This work was supported by the Yahoo InMind project, The Stu Card Fellowship, A Google Faculty Research Award, NSF Grant No. DGE-0903659 and ONR N66001-12-C-4196. The authors thank Eliane Wiese for helping strengthen the message of this paper.

REFERENCES

1. Ackerman, J.M., Kenrick, D.T., and Schaller, M. Is friendship akin to kinship? *Evolution and Human Behavior 28*, 5 (2007).
2. Allen, T. Managing the flow of scientific and technological information. 1966.
3. Backstrom, L. and Kleinberg, J. Romantic partnerships and the dispersion of social ties. *In Proc. CSCW '14*.
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. *In Proc. WWW '10*.
6. Christakis, N.A. and Fowler, J.H. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. 2009.
7. Conti, M., Passarella, A., and Pezzoni, F. A model for the generation of social network graphs. *In Proc. WWMMN '11*.

8. Cranshaw, J., Toch, E., and Hong, J. Bridging the Gap Between Physical Location and Online Social Networks. *In Proc. Ubicomp '10*, .
9. Cummings, J.N., Lee, J.B., and Kraut, R. Communication technology and friends during the transition from high school to college. In *Computers, phones, and the Internet: Domesticating information technology*. 2006.
10. Danezis, G. Inferring Privacy Policies for Social Networking Services. *AISec*, (2009).
11. Eagle, N., Pentland, A., and Lazer, D. Inferring Social Network Structure using Mobile Phone Data. *PNAS 106*, 36 (2009).
12. Eagle, N. and Pentland, A. Social serendipity: Mobilizing social software. *IEEE Pervasive Computing 4*, 2 (2005), 28–34.
13. Festinger, L. Informal social communication. *PsychRev 57*, 5 (1950), 271–282.
14. Gilbert, E. and Karahalios, K. Predicting tie strength with social media. (2009), 211–220.
15. Gliem, J.A. and Gliem, R.R. Calculating, Interpreting, And Reporting Cronbach's Alpha Reliability Coefficient For Likert-Type Scales. (2003).
16. Granovetter, M. The strength of weak ties. *The American Journal of Sociology,*, (1973).
17. Granovetter, M. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, (1983).
18. 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.
19. Hill, R.A. and Dunbar, R.I.M. Social network size in humans. *Human Nature 14*, (2003), 53–72.
20. Kim, J. and Zimmerman, J. Cherish. *In Proc. CHI EA '06*, (2006), 953.
21. Kim, T.H.-J., Yamada, A., Gligor, V., Hong, J., and Perrig, A. RelationGram: Tie-Strength Visualization for User-Controlled Online Identity Authentication. *Financial Cryptography and Data Security*, (2013).
22. Marin, A. and Hampton, K.N. Simplifying the Personal Network Name Generator. *Field Methods 19*, 2007.
23. Marsden, P. V and Campbell, K.E. Measuring tie strength. *Social Forces 63*, 1984, 482–501.
24. McCarty, C. Structure in personal networks. *Journal of Social Structure 3*, 1 (2005).
25. Mesch, G.S. Social context and communication channels choice among adolescents. Computers. *Human Behavior 25*, (2009), 244–251.
26. Min, J.-K., Wiese, J., Hong, J.I., and Zimmerman, J. Mining smartphone data to classify life-facets of social relationships. *In Proc. CSCW '13*.
27. Miritello, G., Moro, E., Lara, R., et al. Time as a limited resource: Communication strategy in mobile phone networks. *Social Networks*, (2013).
28. Neustaedter, C., Brush, A.J.B., Smith, M.A., and Fisher, D. The Social Network and Relationship Finder: Social Sorting for Email Triage. *In Proc. CEAS '05*.
29. Newcomb, T. *The acquaintance process*. New York: Holt, Rinehart, & Winston, 1961.
30. Odom, W., Zimmerman, J., and Forlizzi, J. Teenagers and Their Virtual Possessions: Design Opportunities and Issues. *Proc. of CHI '11*, .
31. Onnela, J.-P., Saramäki, J., Hyvönen, J., et al. Structure and tie strengths in mobile communication networks. *PNAS*, (2007).
32. Pentland, A. *Honest Signals: How They Shape Our World (Bradford Books)*. The MIT Press, 2008.
33. Ricken, S.T., Schuler, R.P., Grandhi, S.A., and Jones, Q. TellUsWho: Guided Social Network Data Collection. *HICSS '10*.
34. 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*, (2009).
35. Roberts, S.G.B. and Dunbar, R.I.M. Communication in social networks: Effects of kinship, network size, and emotional closeness. *Personal Relationships*, (2011).
36. Spencer, L. and Pahl, R. Rethinking Friendship: Hidden Solidarities Today. 2006.
37. Wang, D., Pedreschi, D., Song, C., Giannotti, F., and Barabasi, A.-L. Human mobility, social ties, and link prediction. *In Proc. KDD '11*.
38. Wiese, J., Hong, J.I., and Zimmerman, J. Challenges and opportunities in data mining contact lists for inferring relationships. *In Proc. Ubicomp '14*.
39. 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. *In Proc. UbiComp '11*.
40. Xiang, R., Neville, J., and Rogati, M. Modeling relationship strength in online social networks. *In Proc. WWW '10*, .
41. Zhou, W.X., Sornette, D., Hill, R.A., and Dunbar, R.I.M. Discrete hierarchical organization of social group sizes. *In Proc RSBS '10*.
42. Zipf, G. *Human behavior and the principle of least effort*. 1949.
43. Weka 3: Data Mining Software in Java.