Are You Close with Me? Are You Nearby? Investigating Social Groups, Closeness, and Willingness to Share

Jason Wiese, Patrick Gage Kelley, Lorrie Faith Cranor,
Laura Dabbish, Jason I. Hong, John Zimmerman
Carnegie Mellon University
{wiese, pkelley, lorrie, dabbish, jasonh, johnz}@cs.cmu.edu

ABSTRACT

As ubiquitous computing becomes increasingly mobile and social, personal information sharing will likely increase in frequency, the variety of friends to share with, and range of information that can be shared. Past work has identified that whom you share with is important for choosing whether or not to share, but little work has explored which features of interpersonal relationships influence sharing. We present the results of a study of 42 participants, who self-report aspects of their relationships with 70 of their friends, including frequency of collocation and communication, closeness, and social group. Participants rated their willingness to share in 21 different scenarios based on information a UbiComp system could provide. Our findings show that (a) self-reported closeness is the strongest indicator of willingness to share, (b) individuals are more likely to share in scenarios with common information (e.g. we are within one mile of each other) than other kinds of scenarios (e.g. my location wherever I am), and (c) frequency of communication predicts both closeness and willingness to share better than frequency of collocation.

Author Keywords

Privacy, social networking, relationships, tie strength

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Human Factors

INTRODUCTION

The increasing intersection between mobile and social computing provides many new ways for people to interact with one another. Evidence of this can be seen in the actions of the more than 250 million users who accessed Facebook from a mobile device as of April 2011 [18]. Mobile access allows users to share thoughts and images as they happen, as well as keep up on the minutiae of their friends' lives while out and about in the world.

An ever-increasing number of mobile applications are making it easier for people to share their current location, activity, and other contextual information across many different social networking applications. Some of these applications share information automatically [27,35,41,42, 43]. For example, Google Latitude is a location-sharing service with over 10 million members [24], and allows people to continuously share their location with friends without requiring them to "check-in." This kind of automatic sharing removes the need for people to explicitly indicate they want to share a piece of private information with specific people in their social network.

These new modes for explicit, implicit, and automated sharing raise new concerns about privacy in terms of (i) accidental sharing of information [1,17], (ii) presentation of different versions of self to different social communities [3,4,22,36,39], and (iii) the ever increasing burden of configuring privacy and sharing policies across more and more applications and services [28,31]. These issues represent a growing problem as well as an opportunity area for ubiquitous computing. We posit that UbiComp products and services could help us manage information sharing if they had a model of the interpersonal relationships between different people, one that is better than the social graph we have today where everyone is simply a "friend."

Numerous researchers have identified the importance of interpersonal relationship in expressing sharing preferences [8,32,43]; however, little work has been done to understand which characteristics of interpersonal relationships most influence different kinds of sharing. Is a person's sex associated with the kinds of information they are willing to share? Does frequency of collocation or frequency of communication correlate with a preference for sharing location information or photos from family vacations? Understanding the most salient and predictive characteristics associated with sharing presents an opportunity for UbiComp: this knowledge enables UbiComp systems to aid in the management of privacy settings (and the converse, i.e. sharing policies).

Our work develops the role of UbiComp in understanding dimensions of interpersonal relationships, and further to use

UbiComp' 11, September 17-21, 2011, Beijing, China.

this as a method for managing personal information sharing. We conducted a study in which 42 participants provided details about their relationship with 70 friends and their willingness to share different kinds of information with each of those friends. We conducted a mixed-model analysis of variance based on this self-reported data to examine associations between relationship context and closeness, and willingness to share personal information.

Our results show that how close somebody feels to another person is a stronger predictor of a willingness to share than social group; that frequency of communication is a better predictor of closeness than frequency of collocation; and that people are more willing to share information when it is seen as an exchange of information between both people. In this paper we provide an overview of related work, details of our study design, our findings, and insights for the development of new UbiComp technologies to help automate or semi-automate configuration of privacy and sharing preferences.

RELATED WORK

Past Work in Privacy and Sharing

There has been extensive work examining sharing and social relationships. Belk's exhaustive review of many disciplines' literature on sharing helped him distinguish sharing as a distinct behavior from commodity-exchange and gift-giving as it does not result in compensation or thanks [2]. He draws out "demand sharing" as a behavior found in most families. For example, parents feed, cloth, and shelter their children with no expectation of compensation or thanks. In addition, Belk distinguishes two sharing motives. When "sharing-in" people share things with people they feel close to or desire to feel closer to, as a way of strengthening this relationship. "Sharing-out" involves interactions with people outside of close social boundaries and is generally more like gift-giving or commodity-exchange.

With respect to past work on privacy and sharing in UbiComp, one popular thread of work focuses on sharing one's location. A field study of the Locaccino location sharing application found 79% accuracy in user-specified rules for disclosing their own location [43]. However, this study also found that individual policies are likely to change over time and that these policies will require maintenance, echoing concerns about the potential pitfalls of static policies [39]. Work by Kelley *et al.* explored several friend-grouping interfaces as an instrument for creating privacy rules [29]. They found that different interfaces for producing friend lists affect both the strategies that users employ in creating groups as well as the makeup of the groups themselves. They also found that the groupings their participants created were inadequate for expressing privacy preferences, even just after creation time. Our work explores grouping as one of several features that may predict sharing. Our findings suggest that free-form grouping may not be the most effective method for expressing sharing preferences.

Several pieces of work have found that when sharing location or other contextual information, the person to be shared with is the most important factor in deciding whether or not to share [8,30,32], and in determining the detail/granularity of the information to be shared [33]. Olson *et al.* [37] conducted a study examining 40 different types of personal information and 19 general social "types" that people might share personal information with. In addition to finding that an individual's willingness to share depends on who they are sharing the information with, they clustered "friends" based on similarity of answers, revealing several distinct groups: family, co-workers, public (e.g. salesmen), and spouse. They also found clusters of information that are shared (or not shared); however, they could not succinctly label these categories.

Building on these findings, our work connects characteristics of social relationships with sharing preferences. In contrast with the studies above, we collect information about specific ongoing relationships the participant is a part of, and preferences for sharing information with the other party in that relationship. Most important is that our participants are rating specific individuals that they know, and that they provide explicit details about the nature of each relationship. See Table 1 for a list of the information we capture.

Past Work in Tie Strength

Tie strength is an area of strong interest for online social network (OSN) researchers. While there is still some discussion on the definition of tie strength, Granovetter proposed four dimensions: amount of time, intimacy, intensity, and reciprocal services [25].

Many studies have shown that the vast majority of interaction on social network sites is with small numbers of strong ties. For example, two recent studies suggest that the average number of friends on Facebook is around 180 [5,23], though many users have many more. However, most people on Facebook only interact regularly with 4 to 6 people [44]. A different study examined people who posted and tagged pictures of each other on Facebook, and found that on average people had 6.6 such "friends" [7].

Gilbert and Karahalios took seven dimensions of tie strength (the four proposed by Granovetter and three more) and predicted with 85% accuracy whether a tie between two people was strong or weak [21]. Note that this work only predicts a binary category (strong or weak tie) and has no prediction for role or group. Using a different approach, Xiang *et al.* [45] developed an unsupervised model that could predict a range of tie strengths based on activities on OSNs, though this was not empirically validated with users' perceptions. A different line of work looks at tie stability, analyzing mobile phone data of 2M people over a year [26]. De Choudhury *et al.* pose an interesting question regarding what exactly is a tie, showing that different definitions in an email corpus led to different kinds of results [12]. Our work captures a simplified version of tie strength expressed as closeness. We use closeness as a feature to predict sharing preference and discuss it as an attribute UbiComp systems could sense. Because of the scope of data we collect, our work allows us to examine closeness in relation to different properties than the work described above.

In summary, rather than investigating new ways of measuring or inferring tie strength, we examined the influence of tie strength on location and activity sharing preferences. We also demonstrate that using a single self-reported measure of closeness as a proxy for tie strength can be effective in predicting sharing preferences.

Friend Grouping, Affiliation, and Social Structure

In his seminal work on impression management, Goffman proposed that people perform different roles in order to manage their presentation of self to different groups [22]. More recent work has explored challenges for managing online self presentation. Binder, Howes, and Sutcliffe found that diversity of an individual's Facebook network predicts online tension, and that this tension is between social spheres (such as "work" and "college friends"), not within them [3]. Spencer and Pahl examine individual social structure in detail, exploring differences between "given" and "chosen" relationships and descriptive characteristics of individual personal networks [40].

Farnham and Churchill describe the situation of having conflicting social spheres as "faceted identity" and found through a survey of 631 respondents that faceted identity is common, that these facets are in some cases incompatible, and that one approach for managing some of these facets is to use different media to communicate with different facets [20]. A subsequent qualitative study identified three main facets, or life modes: family, work, and social. This work proposes dealing with these modes online with the concept of focused sharing, where individuals can select subsets of their social network to share with [38]. They also emphasized the ubiquity of smartphones as an important tool for managing relationships and connectedness in social media. We used these same three main facets in our analysis. Our work also extends this past work by exploring faceted identity as one of the factors that can influence sharing of personal information, in the form of different social groups.

There has also been past work looking at automatically clustering groups of online friends based on interactions on OSNs. Past work found manually organizing one's friends into groups to be a high burden [28]. Some proposed solutions include automatically inferring the social context of information on OSNs using a person's social structure [11], privacy wizards that use active learning for configuring privacy policies on OSNs [19], and network clustering algorithms for organizing people into groups [28]. By capturing explicit friend groupings in addition to a variety of other information, we can characterize individual groups and compare the utility of grouping to other features.

Researchers have begun using sensors to infer information about people and their social networks, using audio data [6] and collocation data [10]. Eagle *et al*, studied and modeled human social structure using data from mobile phones [14,15,16]. They collected data on call logs, Bluetooth devices in proximity, cell tower IDs, application usage, and phone status, and compared this data to self-reports. They found that self-reports of physical proximity were heavily dependent on recency and salience of the interactions, and thus did not always match mobile phone data. They were also able to infer 95% of friendships based on call records and proximity.

Our work explores the value of developing a more enhanced social graph that includes notions of groups and roles. We do this using individual self-reports as ground truth. This approach, contrasted with those above, has two main benefits. First, we can make observations that describe a more complete relationship between individuals instead of their relationships within the context of a specific service. Second, this approach does not require the participation of friends whose relationships are being described, thereby mitigating concerns about adoption, ecological validity, and critical mass. Our work looks broadly at the members of our participants' perceived personal network and at an aggregate picture of each relationship, as opposed to one fragmented by technology.

STUDY ON FEATURES OF RELATIONSHIPS AND SHARING PREFERENCES

The goal of this study was to reveal salient features of interpersonal relationships that predict willingness to share information relevant to UbiComp applications, such as location, proximity to another person, and activity. Specifically, we wanted to understand the association between factors such as collocation frequency, communication frequency, closeness, and social group, with preferences for sharing specific kinds of information. We conducted an online study where participants provided basic demographic information and a list of friends. They then associated each friend with relevant social groups, rated their perception of closeness with each friend (tie strength), and stated a willingness to share information with each individual for 21 different sharing scenarios.

Participants

We recruited participants by posting ads in several nationwide online bulletin boards and through two study recruiting websites. Prospective participants were selected based on several criteria:

Age (20 - 50): We wanted people in different life stages, especially with respect to either having or being a child within an immediate family.

Occupation (non-student): We chose to exclude students because they do not easily allow distinctions between work and school groups.

Social network membership (members of Facebook with at least 50 Facebook friends): This was a source for generating friends' names for the study. Additionally, membership in a social networking site indicates that participants are more likely to want to share information about themselves with people they know, allowing us to observe differences in their sharing preferences (as opposed to a person who does not want to share at all).

Mobile device usage (must have a smartphone): We felt that participants with smartphones were more likely to understand the potential values and risks of our UbiComp scenarios.

Participants were compensated $20 for completing task 1, and $60 for completing tasks 2 and 3 (see below). The data collection took place online and participants were given two weeks to complete all parts of the study.

Method

We engaged participants in three distinct activities:

1. Generating a lists of friends
2. Describing each friend in terms of closeness and affiliation with different groups
3. Stating willingness to share different kinds of information with each friend

Friends Lists

We wanted to ensure that participants would answer questions about friends who varied in social group and in closeness, so we asked participants to provide two lists. The first was a list of people based on several categories, which we derived based on qualitative work on relationships [34,40]. The groups were:

- people you currently live with (5 people maximum)
- immediate family members (5)
- extended family members (10)
- people you work with (10)
- people you are close to (10)
- people you do hobbies or activities with (10).

We instructed participants to avoid duplicates. The second list was a list of all of their Facebook friends. We provided participants with instructions on how to download this information from Facebook.

We included everyone from the first list (typically less than 40). We then randomly sampled from their Facebook friend list to reach 70 total friend names, excluding duplicates and names that the participant did not recognize. From here on we will refer to this list of 70 names as the "friend list."

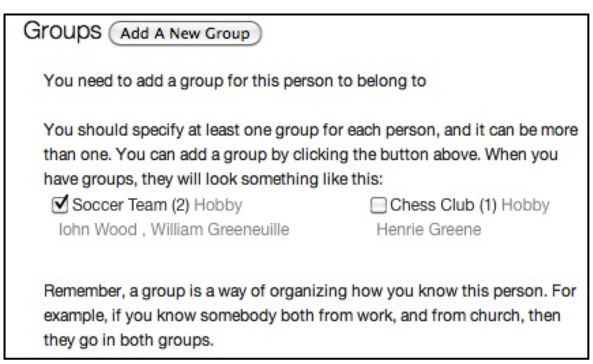

Figure 1: The grouping instructions a participant saw before they had created any groups.

	Data collected	Data type
Observable features	Friend sex	Male/Female
	Friend age	Rounded to nearest year
	Years known	
	Frequency seen	Likert 0-7: Less than yearly (0), yearly, yearly-monthly, monthly, monthly-weekly, weekly, weekly-daily, daily (7)
	Frequency communicated with electronically	
Non-Observable	Closeness (strength of tie)	Likert 1-5: very distant (1), distant, neither distant nor close, close, very close (5)
	Group	Participant-dependent, however each group was put in a pre-specified category

Table 1: Data collected for each friend. Data in the top half ("observable features") is data that we felt was potentially observable by a UbiComp system or social networking site. Data on the bottom half would either be inferred from the observable features or manually inputted by the user

Describing each relationship

We next asked participants to provide information about their relationship with each person on their friend list. The complete list of data collected per friend is in Table 1. We organized this information into two categories: data that we felt would be easily observable from within a UbiComp system or social networking site, and data that would require more work either to infer from observable features or for the user to express manually. Although there are several scales for quantifying tie strength, we opted for a simple measure, asking "How close do you feel to this person?" on a 1-5 Likert scale. This approach is similar to the one taken in work by McCarty [34]. In this paper we refer to this measure as "closeness."

Next, participants detailed their shared affiliations with each friend by placing them into groups that they specified to represent those affiliations. Our interface (see Figure 1) allowed participants to create groups. In addition, it required them to classify each group into one of 12 pre-determined categories: neighborhood, religious, immediate family, extended family, family friend, know through somebody else, work, school, hobby, significant other, trip/travel group, and other. We developed these categories based on a combination of literature sources [34] and data from previous work on grouping friends in social network sites [29]. We instructed participants to indicate at least one group affiliation for each friend, and we encouraged them to indicate multiple group affiliations when relevant. For example, if a person and their friend went to college together, and they both attend or attended the same church, the participant would place them in two groups. The result is a set of affiliations, and all of the people on the friend list who are associated with each affiliation.

n=2370	Mean (SD)	Sharing M=2.83(0.66)							closeness
friend sex = female	55.5%					0.01	0.01	0.01	0.01
friend age	32.7 (12.3)					-0.01***	-0.004***	-0.005***	-0.01***
frequency seen	1.8 (2.2)					0.06***	-0.03	-0.03*	0.21***
frequency comm	2.5 (2.4)					0.17***	0.03**	0.04***	0.34***
years known	10.5 (9.8)					0.03***	0.02	0.01**	0.03***
user age × person age						-0.0004	-0.0004*	-0.0002	0.0001
user sex = female × friend sex = female						0.02	0.01	0.02	0.02
freq seen × freq comm						-0.02***	0.004	0.003	-0.05***
user age × years known						-0.0007**	-0.0004	0.00005	-0.001**
friend closeness	2.7 (1.4)		0.45***		0.40***		0.41***	0.37***	
is family	23.2%			0.59***	0.24***			0.22***	
is social	66.9%			0.14***	0.03			0.03	
is work	18.0%			0.18***	-0.02			-0.001	
Intercept		2.86***	1.62***	3.25***	1.89***	2.33***	1.66***	1.99***	1.61***
R^2 (variance explained)		0.36	0.63	0.48	0.65	0.57	0.65	0.66	0.70
Model Name		User	Close	Mode	Non-Obs	Obs	Obs+close	All	

Table 2: Linear regression models predicting sharing and closeness (last column only), controlling for each participant. Each column is a different model and data in the table are non-standardized β coefficients, except for R^2 in the last row, which can be compared across models to demonstrate the variance explained. For example, the "close" model (fourth column) includes one effect, friend closeness, and this model accounts for 63% of the variance in sharing preferences. Gray cells indicate effects that were not included for that particular model. The data indicate both that closeness is the best predictor of sharing, and that observable features can predict closeness. Significance: *p<0.05; **p<0.01; ***p<0.001

Sharing scenarios

Next, we asked participants to indicate their willingness to share information with each friend in the context of 21 different information-sharing scenarios (see Table 3).

We developed these scenarios using a brainstorming process followed by a broad survey on scenario similarity. First, we brainstormed over 100 different UbiComp scenarios in which individuals could share information, such as location, activity, calendar, history, photos, etc. We grouped scenarios into 11 categories based on the type of information being shared. We assembled these scenarios in a survey that we posted on Amazon's Mechanical Turk, where we asked two questions for each scenario:

- how often do you currently share this information now (whether with one person or with many people): never, seldom, sometimes, frequently, constantly
- how useful is it to you to share this information with somebody you know, answering for maximum usefulness: totally useless, somewhat useless, neither useless nor useful, somewhat useful, totally useful

Using the results from this survey as a guide, we reduced the list of 100 scenarios down to 21 specific scenarios. The resulting list fit into five different categories: current personal location (7), personal location history (5), calendar and location plans (7), communication activity (1), and social graph information (1). See Table 3 for a list of the final set of scenarios used.

For each of the 21 scenarios, we asked participants to indicate their willingness to share information with each of their 70 friends using a 5-point Likert scale (labels: 1-definitely not, 3-no preference, 5-definitely). We adapted this method based on past work by Olson *et al.* [37].

FINDINGS

Forty-two participants completed our study. Their occupations ranged from education and engineering to administration and legal. We eliminated three problematic respondents who each demonstrated no variance for more than 65 out of the 70 friends; each individual friend had the same rating for each of the sharing scenarios. These participants seemed to have simply rated the sharing scenarios as quickly as possible. Of our remaining 39 participants, there were 28 female and 11 male, with ages ranging from 21 to 49 (M=29.8, SD=6.4).

Willingness to Share Location and Activity Information

We were first interested in whether the relationship characteristics that we captured were associated with a willingness to share. Participants were distributed in their mean sharing answer, representing a range of individual privacy/sharing preferences (M = 2.83 out of 5 where 5 is "definitely willing to share this information with this person", SD = 0.66). To address this question we conducted a mixed-model analysis of variance predicting sharing as the outcome variable (see Table 2, note that the variables user age and user sex refers to our study participants). We chose this analysis to account for the non-independence of observations within each participant. This analysis allowed us to explain and compare the variation in sharing using different combinations of independent variables. The goal in doing this analysis was to identify which characteristics of a relationship were most effective for predicting sharing.

All following regressions were done on a per-friend level of analysis; for these models, we took the mean sharing value across all scenarios for each friend (n=2730) and used features that described each relationship as effects in the models. We included the participant in the model as a random effect to account for non-independence of ratings within each participant. The first column shows means and standard deviations for all continuous effects in the model, and percentages of women for the sex data type.

The next column (model name = user) is a model that has no effects except for the effect of the participant (which accounts for individual differences). The result shows that certainly some amount of the variance relates to individual differences, likely preferences for sharing in general (R^2 = 0.36). We also examined models for participant-level effects of sex and age, but the models performed poorly.

Non-Observable Features

The next three columns are models with effects that only include the non-observable data. For these analyses, we pooled together group categories into the three descriptive "life modes" identified by Ozenc and Farnham, (family, work, and social) [38], which they suggest are the primary areas of a person's life.

Closeness by itself turns out to be a very strong predictor of sharing preferences (model name = close, R^2 = 0.63) with each 1-point gain in closeness accounting for a 10% increase of the sharing outcome. This means that a friend who is at closeness 5 (top closeness) is 50% more likely to be shared with. The regression that only had life modes as a predictor did not account for as much of the variance as closeness alone did (model name = mode, R^2 = 0.48), with membership in family, work, and social modes accounting for a 12%, 3%, and 3% increase in likelihood to share respectively (note that all friends were categorized into at least one of these modes). This means that just knowing that somebody is in one of these categories is not particularly helpful in expressing sharing preferences. Finally, adding groups to closeness resulted in only a slight increase in performance over just closeness (model name = non obs, R^2 = 0.65), and resulted in a loss of significance for the "not social" and "not work" effects: closeness and family were all that mattered in this model.

Observable Features

Next, we wanted to explore how well the different observable features (see Table 1) of the relationship predicted sharing, including friend age, sex, years known, frequency seen, and frequency communicated with. Again, we call these observable because we felt that UbiComp systems could capture these features by gathering them from existing social network data, or capturing them automatically using sensor and communication logs. As such, by testing these features, we can evaluate how well a fully automated system might perform for predicting sharing preferences. This model performed well (model name = obs, R^2= 0.57), though still not as well as the model with just closeness. Significant effects included friend age (0.2% less likely to share per year), frequency seen (1.4% more likely to share per point increase), frequency communicated with (3.6% more likely to share per point increase), years known (0.6% increase per year known). The only feature that was not predictive was friend sex.

We included four interactions in the model as well. First, we included the interaction between participant and friend sex and the interaction between participant and friend age to see if homophily accounted for sharing preferences (are men more likely to share with men or women with women, and are people more likely to share with people who are closer in age?), but neither of these were significant.

The next interaction we included was between years known and participant age, which we included because we hypothesized that the duration of a person's life that they have know another person might be a useful indicator. This did have a very small effect, indicating that younger participants were more greatly influenced by how long this person had known them.

Finally, we included an interaction between frequency seen and frequency communicated with. We hypothesized that there are people that are important to us who we communicate with much more often than we see (e.g. family who do not live nearby); similarly, that there are people who are less important to us that we see often but do not exchange as much communication (coworkers who you see often, but with whom you otherwise do not communicate). This interaction was also significant, revealing that communication is a stronger indicator of willingness to share when collocation is less frequent.

Observables and Non-Observables

Next, we wanted to look at models that include observables and non-observables. First we looked at the same model as the previous one but with the addition of closeness. This model improves the explained variance (model name = obs+close, R^2 = 0.65). Closeness has nearly the same effect as in the closeness only model, with each point in closeness increasing the likelihood to share by 8.8%. Frequency seen is no longer significant in this model, neither is the interaction between frequency seen and frequency communicated with. Additionally, frequency communicated with has less of an effect in the model (0.8% more likely to share per point increase, down from 3.6%).

Finally, we included all effects in the model. There was no difference in the variance in sharing behavior (R^2= 0.66), and the model behaved nearly identical to the previous model. As a check, we did run a model with all 12 group categories instead of being grouped into the 3 life modes, and the model did not differ (R^2= 0.67).

Overall, the models with closeness did better than any of the models without closeness, and adding closeness results in the loss of significance for other effects in the model.

Predicting Closeness using Observables

With closeness being such a predictive feature, we wanted to examine how well the observable features of each relationship explain closeness. We used the same approach as before of a mixed-model analysis of variance controlling for participant as a random effect, but this time with closeness as the outcome. We included all observable effects from the other models. This model was quite

Scenario	Pearson's r with closeness	Mean Sharing	Std Dev	Tukey-Kramer HSD			
The next calendar event that we have in common	0.39	3.45	1.42	A			
All calendar events that we have in common	0.39	3.40	1.42	A			
I am with a person who we both know	0.43	3.36	1.39	A			
I'm within 1 mile of this person	0.49	3.26	1.46		B	C	
Details of who my family connections/family relationships are	0.46	3.17	1.39		B	C	D
My personal travel plans that mean we will be in the same place	0.43	3.17	1.54			C	D
My location when I am closer to this person than we normally are	0.35	3.06	1.60			C	D
Everywhere I have travelled to	0.47	3.02	1.36	E			D
My location when I am on vacation	0.53	3.02	1.38	E	F		
I've been to the place that this person currently is	0.42	2.94	1.43	E	F		
My work travel plans that mean we will be in the same place	0.36	2.94	1.62		F	G	
All places that I've been to that this person has also been to	0.42	2.92	1.42		F	G	H
My location when this person has been here before	0.41	2.84	1.37	I		G	H
Everywhere that I have gone out to eat	0.38	2.80	1.34	I			H
I'm at home during a normal weekend	0.50	2.71	1.31	I			
When I am usually at work	0.40	2.45	1.29		J		
My location wherever I am	0.39	2.39	1.34		J		
My tentative plan for the day	0.36	2.23	1.24			K	
I'm in a call on my cell phone	0.25	2.19	1.30			K	L
When the next thing on my calendar starts	0.33	2.10	1.18				L
All details of the next event on my personal calendar	0.33	1.93	1.19	M			

Table 3: Summary of data for each sharing scenario, sorted by overall mean sharing. The first column reports the correlation with closeness, and all correlation coefficients are significant to p<.001. The Tukey-Kramer test compares the overall means for sharing in each scenario: scenarios that have the same letter are not significantly different from each other.

effective (R^2 = 0.70, last column of Table 2). Significant effects in this model included: friend age (0.2% less close per year), frequency seen (4.2% closer per point increase), frequency communicated with (6.8% closer per point increase), years known (0.6% closer per year). The interaction between frequency seen and frequency communicated was also significant, showing that communication has a much stronger effect when collocation is infrequent. The interaction between participant age and years known was significant with a small effect as before. Friend sex and the interactions of participant and friend age and participant and friend sex were not significant.

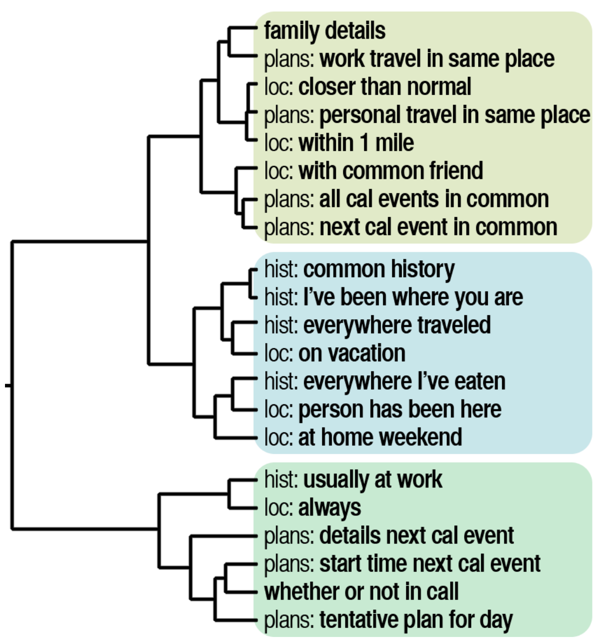

Figure 2: Hierarchical clustering using average linkage distance. Horizontal position of the branches is directly proportional to the calculated distance between each cluster. Scenarios are shorthand for the same ones in Table 3.

Sharing Across Different Scenarios

The finding that closeness was such a strong predictor for sharing across all scenarios led us to further investigate how closeness related to the different scenarios. That is, is closeness a strong predictor for certain scenarios only, or for all scenarios? Our analysis indicates that closeness is correlated with sharing for all scenarios with Pearson's correlation values ranging from r=0.25 to r=0.53, all p<0.001 (see Table 3 for all values).

By asking about sharing across 21 different scenarios, we were able to investigate differences in sharing as a function of scenario type. All scenarios were significantly and positively correlated with each other (r=0.40 to 0.96, Cronbach's α = 0.97).

We examined these similarities further by performing a hierarchical cluster analysis using the average linkage distance formula, a standard technique for examining groupings among items which Olson *et al.* also used in their analysis of privacy and sharing [37]. We chose to use mean sharing per level of closeness as the input because of the strength of closeness in explaining the variance of sharing responses. The dendrogram in Figure 2 shows the clusters. The horizontal scale for the dendrogram is linearly related to the cluster distance at each point where a pair of clusters was merged. For example, in the middle of the dendrogram "hist:common hist" and "hist:I've been where you are" were more closely clustered than the next two "hist:everywhere traveled" and "loc:on vacation", which you can tell by the fact that the first cluster is formed closer to the right side than the second one. Note that the scenarios are shorthand for the same scenarios in Table 3.

These clusters demonstrate an interesting grouping. If we consider the scenarios split into three clusters, the clusters can be roughly labeled as scenarios based on sharing using information about something that we have in common (see Figure 2 top, e.g. loc:within 1 mile); location-history-related scenarios (see Figure 2 middle, e.g. hist:everywhere traveled); and scenarios that reveal sensitive information (see Figure 2 bottom, e.g. loc:always).

To ensure that the means were in fact significantly different, we performed a Tukey-Kramer HSD across all of the means (see Table 3, there was no significant difference across scenarios that are connected by the same letter). This revealed 13 groups (some of which overlap) of scenarios with no mean difference. From the table, we can see that the seven highest-mean sharing scenarios all involve sharing personal information that has something in common with the friend's information, for example shared calendar events or location proximity with the friend.

DISCUSSION

Closeness for expressing sharing preferences

A main focus of our work was to understand which of the collected features are most useful for predicting individual sharing preferences. Our results show that our simple 1-5 Likert scale for closeness was clearly the most useful feature for predicting sharing, outperforming grouping and all other models that do not include closeness.

An examination of the literature indicates that there has been much emphasis on privacy controls that focus on grouping [11,19,28]. In addition, commercial OSNs all seem to either provide grouping controls (e.g. Facebook and LinkedIn), or else require users to specify sharing preferences on a per-friend basis (e.g. Google Latitude's "send my location" feature). We also cannot find a single example of a system (location-based or otherwise) that lets users express closeness for their friends. Not asking about closeness may be a missed opportunity as it is a simple question to ask and is quite predictive of sharing. While it is true that a grouping paradigm does not prevent individuals from constructing groups based on closeness, on the surface it seems easier for users to simply express closeness.

An advantage of using closeness to aid privacy controls is that expressing closeness also expresses a sense of order. That is to say, with group-based privacy controls there is no natural ordering between groups; they are nominal. The ordinal nature of closeness can be useful for expressing privacy controls, as users could simply express "don't share with anybody below medium closeness" (closeness = 3). We could also imagine tiered rules, like "closest friends (5) can always see my location, medium-close friends (3 and 4) can only check up to twice a day, nobody else (1 and 2) can see it without requesting."

To be clear, we are not making the argument that friend grouping is not useful and that researchers and system designers should abandon their focus on this structure. We can imagine systems and types of information that could benefit from an accurate description of groups. In addition, there are many sharing scenarios that we did not ask about in this study (e.g. pictures, status updates, etc), and we did not ask participants about sharing a specific piece of information that might have a more singular group focus. Instead we had them rate classes of information to share. Nevertheless, we do argue that systems would benefit in allowing users to express closeness.

Given our results, it seems that researchers and system designers might find it valuable to investigate automated or semi-automated approaches to inferring closeness.

Sharing scenarios and information "in common"

The scenarios that triggered the highest overall willingness to share all involve an exchange of information that the user and their friend have in common, referred to here as "in-common information." For example, in the 3rd highest ranked scenario (I am with a person who we both know), and the 4th highest (I'm within 1 mile of this person), the sharing scenarios are dependent on the in-common information: we both know this person in the former case and we're within 1 mile of each other in the latter. Compare this to "My location wherever I am," for example, where the sharing is not dependent on in-common information.

First, consider the differences between the clusters in the three-cluster set in Figure 2. The top cluster is mostly focused on in-common information that addresses current or future information (plans). The middle cluster deals mostly with scenarios that have in-common information and a location history component, or scenarios that do not involve in-common information. The bottom cluster does not involve any in-common information, and the scenarios represent less constrained sharing, or scenarios that reveal more sensitive information.

There are several reasons that might explain why scenarios with in-common information resulted in an increased willingness to share. First, people share in order to initiate, maintain, and strengthen relationships with others within their social boundary [2]. Sharing in our scenarios is also a means of self-disclosure, and past work on self-disclosure has revealed the importance of symmetry or equity of disclosure in friendships [9]. These studies have shown that people try to compensate for inequity in sharing practices: if person A demonstrates higher levels of disclosure, person B is likely to disclose more as a result. Similarly, if person A discloses less than person B, person B is likely to disclose less as a result. If we look at the sharing scenarios through this equity-in-self-disclosure lens, perhaps it is the symmetry that motivates the willingness to share.

Another explanation is that people want to maintain privacy (control of their self-disclosures) [13], and that there is some aspect of in-common information that is useful for this purpose. Possible explanations are:

- **utility:** people are more willing to share common information because the information is more useful
- **frequency:** people are more willing to share in these situations because they were less likely to occur
- **no new information:** people are more likely to share in these situations because they are revealing information that the friend might already know

Though these factors are by no means mutually exclusive, it would be valuable to understand which qualities (these or others) make sharing in-common information a more attractive paradigm than sharing other kinds of information. This presents a significant opportunity for future work.

The immediate implication is that there are important considerations for the design of sharing control mechanisms: choosing to support sharing controls that specify in-common information may encourage sharing.

This finding also motivates further work to develop secure and reliable methods for detecting in-common information and only revealing the information (and the associated identities) when particular information does in fact overlap.

Collocation and Communication Frequency

While collocation is clearly a type of in-common information, it did not predict sharing or closeness nearly as well as frequency of communication. This finding seems to support work by Cranshaw *et al.,* which reported that collocation alone was not a useful feature in predicting friendship between two individuals [10]. They found that other features of the collocation history, such as the entropy of a location, significantly improved predictions. Our finding also suggests further work is necessary to determine what features of collocation and communication correspond to features of relationship and to sharing preferences.

One possible explanation for why communication frequency outperforms collocation is that people cannot always choose whom they are near. People are frequently near people that they know, be it coworkers, commuters, acquaintances, or classmates, all of whom they may not feel close to. On the other hand, frequency of communication seems to support more control, especially for the initiator of the communication. Again, there may be communication with work colleagues or others with whom a person is not close. We believe it may be possible for UbiComp systems to differentiate these kinds of communication, for example based on time of day or location, to add richer structure in inferring the nature of the communication, and ultimately the nature of the relationship and closeness.

Limitations

It is important to acknowledge that this data is entirely self-reported, and that further work is required to demonstrate the real-world application of these findings. By conducting the study online and anonymously, we expect that experimenter effects were minimized. Furthermore, individuals are the ground truth on measures such as felt closeness. However, some of the answers may have been idealized responses (e.g. people they call less frequently than reported), or participants may have been unable to answer (e.g. cannot answer for all places I've been to).

CONCLUSION

Social conventions around sharing are being strongly affected by the proliferation of mobile UbiComp technologies, in terms of what information can be shared, how we share it, and how we consume it. With this, the burden of managing one's sharing behavior with the variety of social relationships that one maintains is greater than before, and the systems that we design have the ability to either improve this, or to exacerbate the situation.

This work is a step towards improving that situation. In our study, we found that of all of the data we collected describing 2730 social relationships, self-reported closeness was the best indicator of whether or not to share a piece of information, and common information is more likely to be shared. Also, frequency of communication performs better than frequency of collocation in predicting whether or not to share and in predicting closeness. These results have implications for the design of systems to improve the experience of sharing, and show promise that systems could automatically or semi-automatically predict useful defaults for individual sharing preferences. By automating the process of expressing sharing preferences, we can reduce the burden that systems impose on users and increase the utility of these systems, empowering the user to share information with those whom they would like to share, and otherwise to maintain their privacy.

ACKNOWLEDGMENTS

This work was supported in part by a Yahoo Key Scientific Challenge Award, Google, The Alfred P. Sloan Foundation, and NSF DGE-0903659. The authors wish to thank Ian Li and Rebecca Balebako for their assistance in this work.

REFERENCES

1. Adams, P. The Real Life Social Network. *Voices that Matter Web Design Conference*, 2010. http://www.slideshare.net/padday/the-real-life-social-network-v2.
2. Belk, R. Sharing. *Journal of Consumer Research 36*, 5 (2010), 715-734.
3. Binder, J., Howes, A., and Sutcliffe, A. The problem of conflicting social spheres: effects of network structure on experienced tension in social network sites. *In Proc. CHI '09*, (2009), 965-974.
4. boyd, d.m. and Ellison, N.B. Social Network Sites: Definition, History, and Scholarship. *JCMC 13*, 1 (2008)
5. Burke, M., Marlow, C., and Lento, T. Social network activity and social well-being. *In Proc CHI '10* (2010)
6. Choudhury, T. and Basu, S. Modeling conversational dynamics as a mixed-memory markov process. *In Proc. NIPS '04*, (2004).
7. Christakis, N.A. and Fowler, J.H. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. 2009.

8. Consolvo, S., Smith, I.E., Matthews, T., LaMarca, A., Tabert, J., and Powledge, P. Location Disclosure to Social Relations: Why, When, & What People Want to Share. *In Proc. CHI '05*
9. Cozby, P., Self-disclosure: A literature review. Psychological Bulletin, Vol 79(2), Feb 1973, 73-91
10. Cranshaw, J., Toch, E., and Hong, J. Bridging the Gap Between Physical Location and Online Social Networks. *In Proc. Ubicomp '10*
11. Danezis, G. Inferring Privacy Policies for Social Networking Services. *In Proc. AISec '09*
12. De Choudhury, M., Mason, W.A., Hofman, J.M., and Watts, D.J. Inferring relevant social networks from interpersonal communication. *Proc. WWW '10*, 2010
13. Derlega, V.J. and Chaikin, A.L. Privacy and Self-Disclosure in Social Relationships. *Journal of Social Issues 33*, 3 (1977), 102-115.
14. 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.
15. Eagle, N. and Pentland, A. Eigenbehaviors: Identifying Structure in Routine. *Behavioral Ecology and Sociobiology 63*, 7 (2009), 1057-1066.
16. Eagle, N. Mobile Phones as Social Sensors. *The Handbook of Emergent Technologies in Social Research*, (2010).
17. Elgan, M. The five stages of Facebook grief. *Computerworld*, 2010. http://www.computerworld.com/s/article/9179258/The_five_stages_of_Facebook_grief.
18. Facebook. Statistics. http://www.facebook.com/press/info.php?statistics.
19. Fang, L. and LeFevre, K. Privacy wizards for social networking sites. *In Proc. WWW '10*
20. Farnham, S.D. and Churchill, E.F. *Faceted identity, faceted lives*. *In Proc. CSCW '11*
21. Gilbert, E. and Karahalios, K. Predicting tie strength with social media. *In Proc. CHI '09*
22. Goffman, E. *The Presentation of Self in Everyday Life*. Doubleday, 1959.
23. Golder, S.A., Wilkinson, D., and Huberman, B.A. Rhythms of Social Interaction: Messaging within a Massive Online Network. *In Proc. C&T '07*
24. Google. Check in with Google Latitude. http://googlemobile.blogspot.com/2011/02/check-in-with-google-latitude.html.
25. Granovetter, M. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory 1*, (1983), 201-233.
26. Hidalgo, C.A. and Rodriguez-Sickert, C. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications 387*, 2008
27. Hsieh, G., Tang, K.P., and Hong, J.I. The Design and Evaluation of Privacy Controls and Feedback Mechanisms for Contextual Instant Messaging. (2007)
28. Jones, S. and O'Neill, E. Feasibility of Structural Network Clustering for Group-Based Privacy Control in Social Networks. *In Proc. SOUPS '10*
29. Kelley, P.G., Brewer, R., Mayer, Y., Cranor, L.F., Sadeh, N. An Investigation into Facebook Friend Grouping. *In Proc. INTERACT '11.*
30. Khalil, A. and Connelly, K. Context-aware telephony: privacy preferences & sharing patterns. *Proc. CHI '06*
31. Lampinen, A., Tamminen, S., and Oulasvirta, A. All My People Right Here, Right Now: management of group co-presence on a social networking site. *In Proc. CSCW '09*
32. Lederer, S., Mankoff, J.C., and Dey, A.K. Who Wants to Know What When? *In Proc. CHI '03*
33. Lin, J., Xiang, G., Hong, J.I., and Sadeh, N. Modeling People's Place Naming Preferences in Location Sharing. *In Proc. Ubicomp '10*
34. McCarty, C. Structure in Personal Networks. *Journal of Social Structure 3*, (2005).
35. Miluzzo, E., Lane, N.D., Fodor, K., et al. Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the CenceMe Application. *In Proc. SenSys '08*
36. Odom, W., Zimmerman, J., and Forlizzi, J. Teenagers and Their Virtual Possessions: Design Opportunities and Issues. *In Proc. CHI '11*
37. Olson, J., Grudin, J., and Horvitz, E. A study of preferences for sharing and privacy. *In Proc. CHI '05*
38. Ozenc, F.K. and Farnham, S.D. Life " Modes " in Social Media. *In Proc. of CHI '11*
39. Palen, L. and Dourish, P. Unpacking "Privacy" for a Networked World. *CHI Letters 5*, 1 (2003), 129-136.
40. Spencer, L. and Pahl, R.E. *Rethinking friendship: hidden solidarities today*. Princeton University Press, 2006.
41. Tang, J.C., Yankelovich, N., Begole, J., Kleek, M.V., Li, F., and Bhalodia, J. ConNexus to awarenex: extending awareness to mobile users. *In Proc.CHI '01*
42. Tang, K.P., Keyani, P., Fogarty, J., and Hong, J.I. Putting people in their place: an anonymous and privacy-sensitive approach to collecting sensed data in location-based applications. *In Proc. CHI '06*
43. Tsai, J., Kelley, P., Drielsma, P.H., Cranor, L.F., Hong, J., and Sadeh, N. Who's viewed you? the impact of feedback in a mobile-location system. *In Proc. CHI '09*
44. Wilson, C., Boe, B., Sala, A., Puttaswamy, K.P.N., and Zhao, B.Y. User interactions in social networks and their implications. *In Proc. EuroSys 09*.
45. Xiang, R., Neville, J., and Rogati, M. Modeling relationship strength in online social networks. *In Proc. WWW '10*