The Composition and Use of Modern Mobile Phonebooks

Frank Bentley and Ying-Yu Chen
Yahoo Labs
San Francisco, CA USA
{fbentley, yyc}@yahoo-inc.com

ABSTRACT
Over the past decade, the mobile phonebook has evolved from a relatively short list of people that one calls and texts to a many-hundred person list of aggregated contacts from around the web. This is happening at a time when an increasing number of mobile applications are relying on the mobile phonebook to create one's social network in their services. Through a large-scale study of the phonebooks of 200 diverse participants, containing 65,940 contacts, we set out to understand today's mobile contact lists. Our participants reported that they did not recognize the names of 29% of their contacts and we found that the most frequently contacted five contacts represent greater than 80% of all calls and text messages with phonebook contacts. We conclude with implications for the design of mobile applications that rely on phonebook data.

Author Keywords
Mobile; Phonebook; Communication.

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

INTRODUCTION

Mobile phones are becoming the main tool for daily communication needs. Recent research has shown that people are actively using their phones for hours per day and spending the majority of their time in communications-related applications [5]. An increasing number of these applications are relying on the mobile phonebook to create one's social network in their services. For some apps, phonebook entries are seen as "trusted" contacts who all get to automatically know if you're using the same app. In some cases, apps automatically share various rich presence information related to a contact's context.

In parallel, since the introduction of MotoBlur in 2009 [2], mobile phonebooks are now aggregators across many networks (calling, multiple email providers, Facebook, LinkedIn, Twitter, etc.) making them much more than just lists of people that you call on the phone. However, there has not been a systematic study investigating these new social phonebooks since they were first launched publicly five years ago. This is despite over one billion active Android users[1] having access to a variety of different forms of aggregated phonebooks and major applications with hundreds of millions of users accessing these lists. Who is really in your socially-connected phonebook? And what can this say about the suitability of this list for various services (location sharing, auto-friending, etc.)?

We set out to explore these questions for the purposes of improving the design of mobile phonebooks and providing implications for the design of new communications applications that take advantage of mobile phonebook data. Through a study of 200 diverse participants from across the United States, we analyzed phonebook entries and call log data for their 65,940 contacts. Participants reported that they had "no idea" who 29% of their contacts were and we found that the most frequently contacted five contacts represent greater than 80% of all calls and text messages with people in the phonebook. This has strong implications for current phonebooks that attempt to treat an average of 308 contacts equally, and for new communications systems which need to make sense of contact data to onboard users into their services.

RESEARCH QUESTIONS

We began this work with a series of research questions related to understanding the composition of current smartphone contact lists, current communication practices on mobile devices, and how contacts might be used in a variety of applications. With respect to the phonebooks, we wanted to know: How many contacts do people have? Which services are linked? Who are these contacts? Are they current friends/colleagues? Former contacts?

With respect to communication history, we were interested in exploring: What percentage of contacts do people actually communicate with? How often? What percentage of communication goes to the top five or ten contacts? When was last time of communication to specific types contacts in any form?

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea
Copyright 2015 ACM 978-1-4503-3145-6/15/04…$15.00
http://dx.doi.org/10.1145/2702123.2702182

[1] http://www.engadget.com/2014/06/25/google-io-2014-by-the-numbers/

Finally, we were interested in exploring how contacts might be used in applications that integrate with mobile phonebooks. Particularly, we wanted to understand: How comfortable would people be in sharing various aspects of their lives (location, TV viewing, photostream, calendar, etc.) with these contacts? Are there differences between contacts that users call or text regularly and other contacts? How comfortable would people be with automatically adding specific contacts in particular kinds of applications (e.g. chat, location sharing, etc.)?

In order to explore these questions, we designed an online survey that integrated with participants' mobile phonebooks to collect data. We will describe the survey and methods below after introducing some background and related work.

RELATED WORK

Research into mobile phonebooks and communication patterns has been explored since the early 2000s. This research has included developing new phonebooks with a variety of features, studying their use, and studying call logs to understand fundamental qualities of daily life such as commute patterns. However, we could not find an example of a paper that deeply explored the use of the new, socially-aggregated contact lists that are common on over a billion of today's smart phones. Nor could we find a study of current mobile communication practices to phone contacts. This type of information is essential to build new communications systems as well as to understand how communication has changed in the era of the smartphone.

Early work in the field of Mobile HCI explored augmented phonebooks. Context Phonebook [22] was an example of this work that augmented phonebooks with rich presence information about contacts, such as their location or current ringer profile. ContextPhone [20] was a later example of this type of system. Motion Presence from Motorola Labs [4] took another spin on this idea, and provided a more privacy-preserving, binary "in-motion" indicator next to the names of close friends and family in the phonebook to help moderate availability and support micro-coordination [16]. These projects all explored what it would mean to have rich, real-time social information about contacts visible in the phonebook and how people would engage with such systems.

This early research led to commercial phonebooks that aggregated data from social networks. The MotoBlur service [2] became the first commercial social phonebook, aggregating contacts from Facebook, MySpace, Twitter, and other social networks into the phonebook. This necessitated a matching process that would link together contacts from multiple services. It also led to contact lists that ballooned from the small number of people one actually calls and texts to lists of potentially thousands of contacts aggregated from one's "friends" across social networks. This type of phonebook is now the default on Android and Windows Phone and has been largely unstudied since its introduction over five years ago.

A number of researchers have recently explored using call data records (CDRs) from mobile operators to understand patterns of human mobility. Eagle [9] set out a wide-ranging research area to use mobile phone logs to explore large populations and minimize the errors of data that is collected using self-reports. He explored the possibility of using large-scale CDRs for understanding human mobility patterns, social network structure, and the ability to confirm theories such as Milgram's social distance experiment. Furletti *et al.* [10] explored using CDRs to estimate mobility flow in and out of a city. We were inspired by this area of work and explicitly chose to base our communications findings on real call and SMS logs uploaded from phones, instead of relying on self-reported communication frequency.

Several researchers have studied SMS usage patterns in smaller-scale studies. Battestini *et al.* [1] explored SMS behaviors with 70 university students, finding that on average they texted 47 contacts. Grinter and Eldridge [11] studied ten British teenagers and found they they texted between 10-15 other people. Ito [14] studied the phonebooks of Japanese teenagers, finding that while they had many contacts, they only communicated with a few people. We were curious to see what communications practices looked like for a broader audience and with modern smartphones, a combination that we have not seen studied in detail.

Other researchers have explored tie strength with regards to the contacts in mobile phone books. In perhaps the closest study to our work, Min *et al.* [19] explored the contacts in mobile phonebooks and asked 40 participants to rate the tie strength to a curated selection of contacts. However, in their focus on tie strength, the authors did not explore the contact list more fully. In particular, none of our research questions above to understand the larger phonebook and communication patterns can be answered with the data in their paper. Because they did not choose random contacts from the phonebook, the overall makeup of the phonebook cannot be determined. Wiese *et al.* [24] extended this work to show that contacts who are contacted frequently are often high tie-strength, but contacts that are not frequently contacted on the telephone are not necessarily weak ties.

In addition to the academic literature, there are now a large number of commercial mobile applications that upload contact data or otherwise attempt to integrate with the phonebook. Wildly popular applications such as WhatsApp, Facebook, LinkedIn, Twitter, Instagram, Yelp, Path Talk and Foursquare all either automatically, or with user consent, access the mobile phonebook and use the contacts that are stored there to set up lists of potential communication partners [23]. Other smart mobile phonebooks, such as the current Android and Windows Phone contact lists also provide some level of social

integration. The design of these default phonebooks can be enhanced with deeper knowledge of how people communicate and the types of contacts that are stored.

The literature contains gaps in understanding the content of socially-connected mobile phonebooks and current mobile communication patterns. To this end, we created our study to quantify who is in current mobile phonebooks, communication patterns with these people, and the suitability of automatically sharing information with these contacts or adding them in a variety of mobile services. This knowledge will help anyone designing new mobile communication applications or any services that wish to integrate with the mobile phonebook. In addition, we provide an understanding of current communication practices that is interesting from a sociological viewpoint. We will return to this topic in the discussion.

METHOD

In order to answer our research questions, we conducted an online survey in the summer of 2014. We chose to create an Android application that participants could install to collect metadata about their phonebooks and call histories. Android currently has a global market share of 80% of all smartphones [18] and is interesting as it enables complete access to the phonebook from any application with no further user consent than approving the installation of the app. Android phonebooks also merge their contacts from several services into the global contact list (e.g. Gmail, Facebook, LinkedIn). This merging of a variety of different types of contacts into the phonebook has the potential to include a wide set of people, changing the contact list from its traditional role of a relatively short list of people that a person actually calls or text messages. This change has deep implications for the design of mobile applications that rely on phonebook data as discussed above.

Survey

The study was conducted as an online survey, with 200 diverse participants from around the United States. Participants downloaded a small Android application to their phone which retrieved phonebook metadata from the device. We did not collect any information that could be used to contact the people from the phonebook, such as phone numbers or email addresses. Instead, we uploaded the "Name" field (for identification during the survey) as well as boolean values indicating if a phone number or email address was present for each contact. We also uploaded metadata about call and SMS history (timestamp and duration/length only, not actual message content). All information was sent via HTTPS to our server, and each participant was given a code when this upload completed that they could enter into the web-based survey to continue. Participants were shown sample data and had to explicitly agree to the collection of their data before installing the mobile application.

Upon entering the code, we selected 25 random contacts from their uploaded phonebooks and asked participants several questions about these people. They were asked to identify the type of relation that they had with the person (e.g. friend, family, colleague, business, unknown, etc.), the last time they communicated with that person in any way, questions on how comfortable they would feel automatically sharing various types of information with each contact, as well as four questions to assess their tie strength with each contact. For tie strength, we used the questions and scales presented in [24]. These include: "How close do you feel to this person?" "I talk with this person about important matters," "I would be willing to ask this person for a loan of $100 or more," and "I enjoy interacting with this person socially."

Following the questions relating to their 25 randomly selected contacts, we asked for general information about their phonebooks (e.g. which services they synced to their phonebook) as well as general demographic information. The survey took an average of 17 minutes to complete, and participants were compensated for their time.

Participants Demographics

Our 200 participants comprised a very representative sample of the adult American population. They ranged in age from 18-64 (mean = 30, $\sigma = 8$), 51% were female, and 34% had completed college (compared to a US average of 33.5% [21]). They resided in 35 different states, including all regions of the country, and 80% lived in urban areas (compared to a US average of 81% [15]). All had owned their smartphone for at least 6 months. We chose to conclude the study upon reaching 200 participants as the data was not significantly changing after the addition of the final 50 participants. For example, the mean length of the contact lists changed by fewer than 10 contacts (3%).

RESULTS

From our 200 participants, we collected information about 65,940 total contacts. We also received detailed survey responses relating to 4,975 contacts (the random 25 chosen for each participant). A total of 25 contacts (0.5%) were skipped by participants while completing the survey. We also received call logs for 74,493 calls and SMS metadata about 306,228 messages totaling 14,711,628 characters.

In addition, we collected qualitative responses from two open ended questions: one asking participants if anything had surprised them after participating in the study and another asking if they planned on changing anything in their phonebooks after participating. All together, we captured 227 responses that were longer than a single word and analyzed them for themes in an affinity analysis. Themes from these notes will be discussed in parallel with the quantitative data to enhance the presentation of the statistics with explanations from our participants.

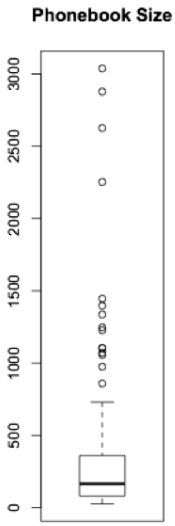

Figure 1: The number of contacts in an Android phonebook. Note that while the mean was 308, there were fourteen participants (7%) with more than 1000 contacts.

In the following subsections, we will describe the composition of today's mobile phone books, explore current mobile communications patterns, and discuss the desirability of sharing various information with particular types of contacts.

Phonebook Composition

Figure 1 shows the distribution of the number of contacts in an Android phonebook. The mean size was 308 contacts with 25% of participants having contact lists with greater than 360 contacts. Phonebooks ranged in size from 26 to 3,038 contacts.

Where did these contacts come from? As expected on a Google Android phone, 84% of participants reported connecting their Google/Gmail accounts to their phonebook. 70% of participants reported connecting their Facebook accounts, while 17% connected their Yahoo accounts. As 71% of American adults use Facebook [7] and Yahoo's mobile email market share is currently 19%, this shows that a high percentage of people choose to sync these services to their phonebook (which is the default behavior on Android for these applications). Eleven percent of participants did not sync their phonebook with any of these services.

Who were these contacts? We asked participants to categorize each of their 25 randomly selected contacts into the categories shown in Figure 2. Most strikingly, over 29% of contact names could not be recognized by the participants (the survey response was phrased "I have no idea"). This is likely due to the fact that a variety of services were linked to their account, including Facebook or LinkedIn where they are likely to have contacts that they have never met in person. This is validated by the open ended responses that were placed at the end of the survey.

Participants were surprised by how many people in the phone book they didn't know or that they didn't communicate with. In the open-ended responses, they shared that they did not previously realize who was in their phonebooks: "I have a lot more people in there that I have no idea who they are than I thought." Another told us: "I realized I have no clue who most people in my contacts are, and yes, it has been VERY surprising!" Several participants even thanked us for showing them how many people they did not know and said that they planned to clean out their phonebooks after participating.

Many participants did not know how contacts got in their phonebook. For example, one participant expressed his/her confusion: "I'm surprised that there are as many people in there that I don't remember adding or calling or texting." Some participants expressed concern as they had no idea how the contacts got in: "I had no idea I had so many contacts!!! I am shocked and a little concerned."

Some participants were surprised that Gmail contacts were in their phonebooks, and reported not liking this. These were often called "junk contacts" or "random contacts," especially those that came from services such as Craigslist. This confused some participants: "I'm confused all the contacts came from. I assume the app used my email too? Because wow, that's a lot of contacts." Some responses

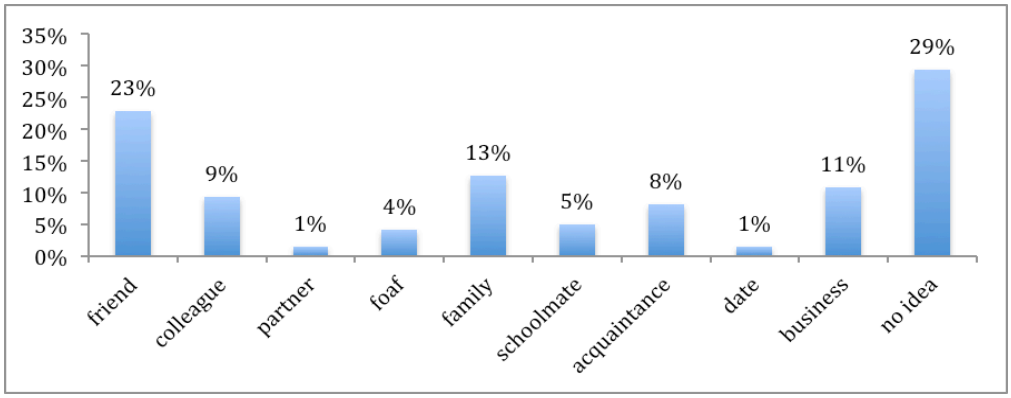

Figure 2: Percentages of contacts by category. There are more contacts that the participants could not identify than friends.

indicated that they did not like this feature of Android. One participant told us, "I really hate how Google automatically adds anyone I ever received email from."

Similar responses came with regards to contacts from other services: "There's lot's of junk there, including Skype contacts," or, "People I follow on twitter are there for some reason." Completing the survey helped them realize that there were contacts in their phonebooks from social networking apps: "I never knew that contacts from social applications (Tango, Viber, Skype, etc.) are included as part of the contacts."

As to the other contacts that participants did recognize, 23% were Friends, 13% were Family, and 9% were Colleagues. We find it interesting that 11% of the contacts were Businesses (such as restaurants, repairmen, etc.). Figure 2 contains more detail on the other categories. Note that a single contact can be in multiple categories.

Turning to tie-strength as a different measure of closeness, 46% of all contacts were rated with a tie-strength of 0 (the lowest possible score on all four questions), as shown in Figure 3. Interestingly, the remainder of the contacts fell fairly evenly between 1-6% for each remaining value of tie strength. We find this encouraging as participants clearly thought about their responses and did not simply answer all four tie-strength questions the same way. It is interesting to note that 6% of contacts were rated as the highest level on all four questions relating to tie strength, representing 10 contacts in the median sized phonebook.

Investigating the contact fields that are populated for each contact, we observed that 55% of phonebook entries have a phone number populated, while 45% have an email address. 5% of contacts have both an email address and a phone number, while another 5% do not have any phone number

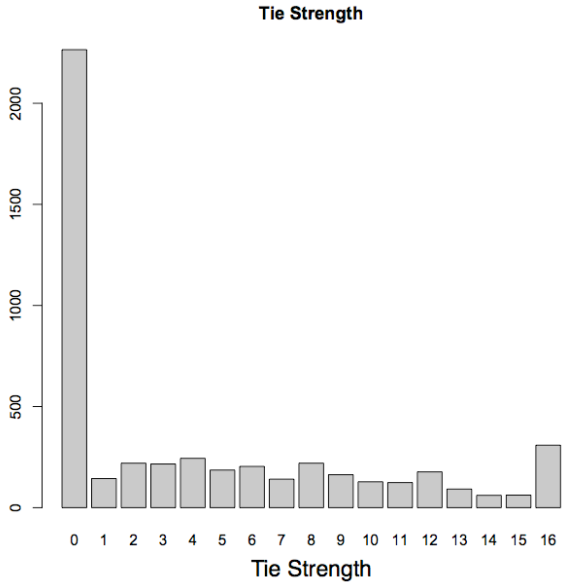

Figure 3: The number of contacts of varying tie strengths. Note that 46% of all contacts have a tie-strength of 0 while 6% have the maximum possible tie strength of 16.

or email address populated. These contacts without contact information can appear from linked sources such as Facebook, LinkedIn, or Twitter where the connected user does not share their phone or email contact information with the participant.

Most contacts conformed to a "FirstName LastName" pattern, however 26% of contact names did not contain a space and are presumed to be first names or nicknames only. A look through these contact names shows examples

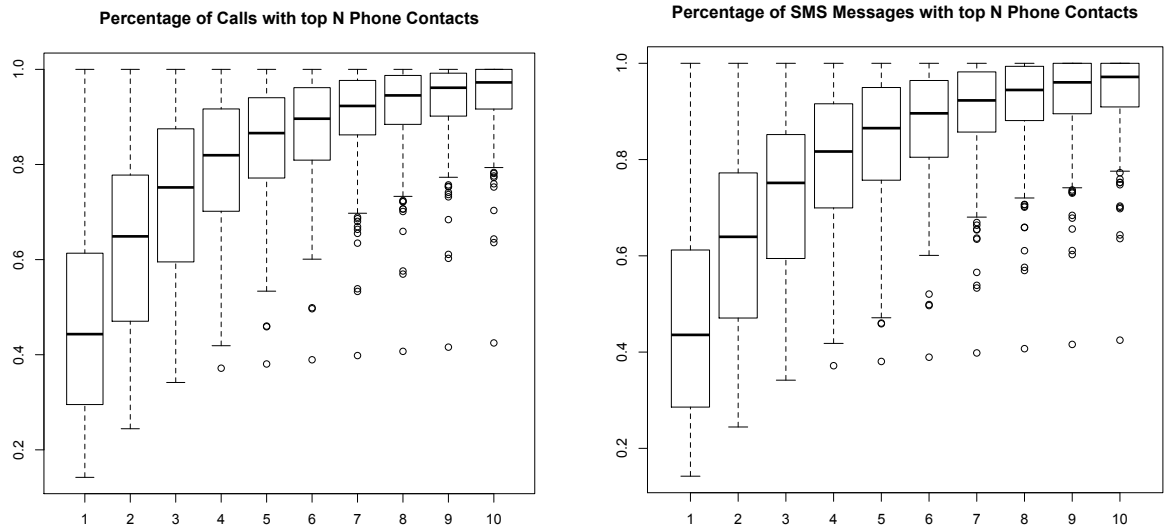

Figure 4: The percentage of calls (left) and SMS messages (right) for the top N most frequently contacted users from the contact list. On average, over 80 percents of all calls and SMS messages are with the top 5 contacts.

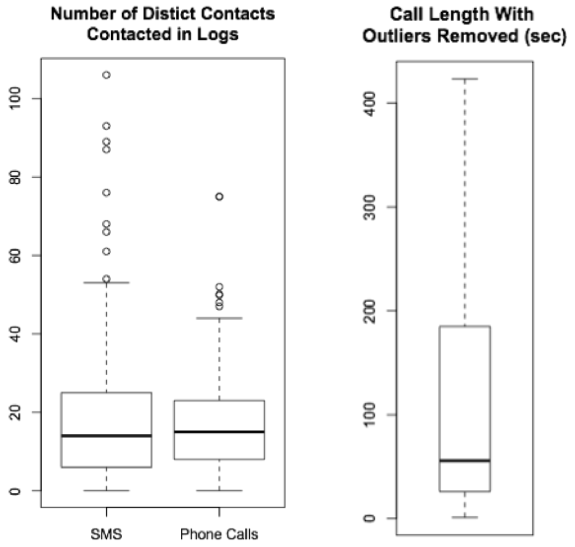

Figure 5: Left: Number of distinct contacts that were contacted via SMS (since phone was purchased) or via phone calls (in the past 500 calls). Right: The length of calls, in seconds. Note that outliers are removed and there were some very long calls (some up to 10 hours) in the dataset.

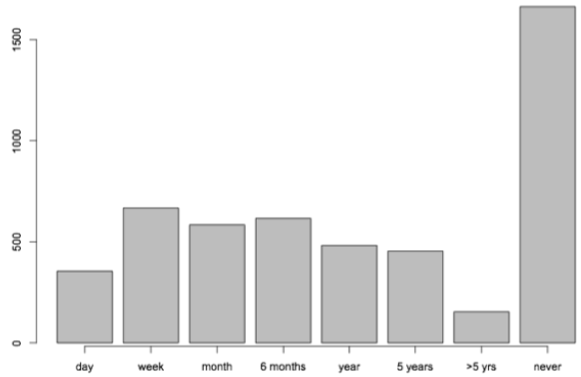

Figure 6: Last time of contact across any medium (including social networking and face to face) for the contacts shown to our participants.

such as "Work," "Bank," "School," "Stinky," and "Roomie." However, even contacts with spaces do not necessarily represent proper names. We saw entries such as "My Wife," "Lover Boy," "Handsome Husband," and "Mom and Dad" in the name fields. This is important for services that attempt to merge contacts with online profiles as name matches are not possible for these types of contacts. It is also important for services that expect to use the contact's name for creating profiles or sending out invitations to join a service. Getting an invite addressed to "Stinky" might not be what that contact would expect or appreciate.

The most common names in the phonebooks were "Mom" (10% of participants), "Dad" (9% of participants), and "Home" (6% of participants). Overall, we can see that mobile phonebooks today contain a wide array of contacts from many broad contexts of a person's life, many of whom are not even recognized by name.

Contact Patterns

After investigating the call logs, we have found that 84% of all calls and 81% of SMS messages with people in the phonebook are to the five most frequently contacted people. Looking at the ten most frequently contacted people yields 94% of phone calls and 92% of all SMS messages that are made with people in the phonebook (see Figure 4).

As the average phonebook size was over 300 entries, we wanted to investigate what percentage of these contacts were ever contacted through phone calls or SMS. Most Android phones limit the call log length to the past 500 calls. Of these 500 calls, an average of only 17 distinct contacts (6% of all contacts) were represented. SMS logs trace back much farther on average to include anyone messaged since the phone was first used. Our participants had an average of 19 distinct contacts (also 6% of all contacts) in the SMS logs. A surprisingly few number of contacts make up the total of all phone or SMS communications, as can be seen in Figure 5. An interesting population to consider is the 25% of participants who had called fewer than eight contacts or messaged fewer than six.

These numbers are similar to the 10-15 contacts that were text messaged in Grinter and Eldridge's study of British teens [11], and much lower than the 47 contacts found in Battestini *et al.*'s college student population [1].

It is also interesting to explore recency of contact. Of contacts that were text messaged, the median length of time since last contact was 26 days with 25% of text messaged contacts having been contacted in the past 8 days. For contacts that were called, the median length since last contact was slightly longer at 36 days with 25% of called contacts being called in the past 17 days.

Digging deeper into the call logs, 27% of all calls were incoming and answered, 19% were incoming and unanswered, and 54% of calls were outgoing calls placed by the participant. We find the high number of missed calls quite interesting, and have not previously seen this reported in the literature. It is also interesting to note that 31% of all calls were with people who were not stored in the phonebook at all. This number was slightly higher for outgoing calls which represented 55% of calls with participants not in the phonebook. It is notable that 65% of all incoming calls from unknown numbers (not stored in the phonebook) were not answered, representing a large majority of all unanswered calls.

If we investigate the self-reported time of last contact in Figure 6, we can see that 33% of all contacts have never

been contacted at all. In total, participants reported that 55% of contacts had not been contacted in any way in over 6 months (including face to face or through social media). This compares to 80% of contacts that had not been contacted in the past six months as seen through the "last contacted" field in the phone logs (which can include services other than calling or messaging). Clearly, these logs are not enough to fully capture the people that one communicates with, since much communication occurs in other applications or face to face and is thus not visible from the phonebook metadata.

As for the communications themselves, phone calls averaged 271 seconds in length. The median call length was 56 seconds, with 75% of calls being shorter than 185 seconds. However, there are some very long multi-hour calls in the logs. Because of these, the call lengths shown in Figure 5 have outliers removed to enhance readability. Just under 1% of calls lasted over one hour and just over 2% of calls lasted between 30-60 minutes. For SMS messages, the average message length was 48 characters and the average message history with an individual contact totaled 75 messages (ranging from 1 to 6645 messages).

Since most mobile phonebooks contain a "recent calls" list, we were interested in investigating how well the recent calls list represented the contacts that are most likely to be contacted. We looked at the most recent N calls for N up to 10 in order to investigate what percentage of the most frequently contacted N contacts are represented in a recent contacts list of length N. The results can be seen in Figure 7. Note that for ten contacts, only a median of six of the most frequently contacted ten contacts appear in this list of the ten most recent calls. While a recents list might be great to contact "temporary" contacts that might be relevant over the course of a day or weekend, it does not provide easy access to the most frequently contacted people.

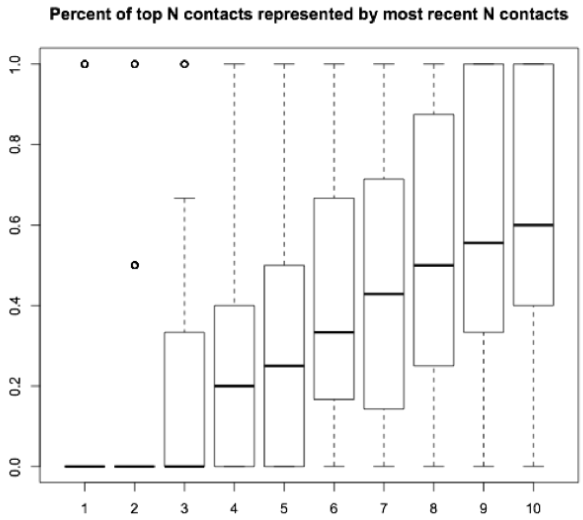

Figure 7: An investigation into the suitability of the "recent calls" list to represent the most contacted contacts. A contact list of the ten most recent calls only captures a median of 6 of the most frequently contacted contacts.

Desirability of Sharing

As the mobile phonebook is now being used by a wide variety of mobile apps to auto-populate contact lists or to determine people to automatically share information with, we were interested in exploring how users felt about sharing a variety of information with their 25 randomly selected contacts. While we recognize that the CHI literature points to complex contextual factors in the decision to share rich presence [6], since many applications with hundreds of millions of users ask each user to choose friends and sharing preferences upon installation, we found these to be relatively important questions to ask.

Table 1 highlights sharing preferences for a variety of

	Friend	Colleague	Partner	FOAF	Family	Schoolmate	Acquaintance	Date	Business
Location	41	16	79	20	46	33	16	31	5
TV Shows	36	13	67	21	40	35	14	28	3
Calendar	32	14	70	16	38	37	11	15	4
Transportation	31	12	71	13	37	30	15	22	4

Table 1: The percentage of contacts from different categories that participants feel comfortable always seeing their location, TV shows they are watching, calendar free/busy times, or mode of transportation.

	Friend	Colleague	Partner	FOAF	Family	Schoolmate	Acquaintance	Date	Business
Chat App	53	22	71	22	51	48	21	34	4
Music App	42	17	64	19	42	41	15	26	3
Location App	31	12	60	10	36	27	9	16	2

Table 2: The percentage of participants who felt comfortable automatically adding a contact as a 'buddy' in a variety of apps.

different presence information. These types of information have been explored in previous research (e.g. Consolvo *et al.* [6], Bentley *et al.* [3], Harboe *et al.* [12]) and are along the lines of information currently being shared in commercial products, for example Facebook's recent addition of TV and Location sharing to its mobile application and Path Talk's sharing of location, music, and mode of transport.

Unsurprisingly, participants felt most comfortable sharing rich presence information with their partner, with 79% willing to share location, and 70% willing to share their calendar. Family, friends, and schoolmates clustered at similar percentages in the 33-45% range while colleagues, acquaintances, and businesses ranked lowest. The differences between the three groups mentioned above are all statistically significant with X^2 between 25-276 and $p<10^{-8}$ for each pairwise comparison between groups. Tie-strength was positively correlated with all aspects of sharing rich presence information and for friending in mobile applications, with $r>0.5$ and $p<10^{-16}$ for all comparisons between overall tie strength and each type of rich presence data.

We find it interesting that our participants did not indicate a significant difference in willingness to share different types of information. The type of person mattered far more than the type of information shared.

We found similar differences when participants were asked if they were willing to let a new application directly connect them to their contacts. Here we also see that participants are more willing to automatically connect to their partners, followed by the set of friends, family and schoolmates, followed everyone else (see Table 2). We will return to the implications of these findings in the discussion.

Actions to Clean Up Their Phonebooks

In the open-ended responses at the end of the survey, many participants indicated that they planned on cleaning up their phonebooks in some way. Some participants decided to unsync specific services. People said that they would "disconnect my Facebook friends from my phone" or "unsync contacts from Google."

Aside from unsyncing entire applications, many participants said that they would remove specific contacts from their phone books. They said that they would "clear out," "delete," or "weed out" people they didn't know or the "unnecessary" contacts. Our participants specifically pointed out "old contacts" as people that they would like to delete. One participant told us: "I will be deleting contacts I no longer talk to because I know for a fact I do not talk to over 200 people. I feel that it is cluttering my phone." Another person told us that they will "remove contacts I no longer utilize on a regular basis (i.e. people I don't call at least once in a year)." Some of these contacts were "one off" contacts such as "people I had classes with at college" and were no longer relevant to the participant.

This idea of "temporary contacts" was interesting to us. For example, many participants had contact information for people that they bought or sold something with on Craigslist. After the transaction was completed, these people were no longer needed, yet almost always remained in the phonebook due to the effort required to find and then delete a specific contact.

Some participants said that they would like to delete some contacts, but that they are too lazy and it takes effort. One participant said that "it's probably going to take more than an hour to do." Another said that they had "a long trip as a passenger tomorrow and will spend some of the time deleting [contacts]." Yet, for others it was just too much effort. They told us: "I don't know if I care to go through that trouble [to clean up the contact list]" or "I'm lazy and it doesn't hurt anything to have them in there." And some participants liked having every contact always available: "I'm not sure I will delete any since I like to keep numbers handy should I ever need them for any reason."

Overall, most users wanted to remove some contacts, but realized that the effort required and the sheer number of contacts that they would have to look through meant that this would be a large task.

DISCUSSION

A thorough analysis of the contents of modern Android phonebooks, which make up 80% of the smartphone market, enables us to explore the implications of this work on the design of new mobile social applications. Particularly, the design of applications that import contact data or new phonebook interactions can benefit from our findings. In addition, our analysis of mobile communication patterns is interesting from a sociological point of view.

Usage Based Favorites

A major finding was that despite having an average of 308 contacts in the phonebook, over 80% of calls and messages were to just five people, and over 92% of communications occurred with just ten people. Despite this, most mobile phonebooks still provide the A-Z list as the most common selection mechanism. Some have the ability to "favorite" a contact, but this feature seems rarely used and is not always very visible for users. Even if contacts are favorited, launching the Contacts app on most Android phones still defaults to the A-Z list.

To better align with user behavior, contact lists (and applications that import these lists) should help users easily find their top ten contacts based on available communication history. All of this can be done on-device without the need to upload contact data to a server and can provide the user with the person they are looking for over 92% of the time without the need to scroll through hundreds or even thousands of names.

"Trusted" Contacts

Applications should not assume that just because a person is present in a mobile phonebook, that they are a trusted contact. Over 29% of all contacts in our participants' phone books were completely unknown to them, and only 6% scored the highest in perceived tie-strength. Applications should see the phonebook as a very extended social graph, including many friends of friends or former acquaintances. While there might be some services where it is appropriate to include all of these people as contacts, our research shows that people are highly uncomfortable automatically sharing location or calendar data with over 80% of the people in their phonebooks. Applications should be careful to allow users to select the specific contacts that will be able to see a particular aspect of their context.

Aggregated Lists

Back when MotoBlur was first designed [2], it was assumed that people would want all of their contacts accessible on their mobile device. However, as the number of social networks that a user belongs to has increased and friend counts on these networks have also increased, it has become clear that for some users there are just too many contacts in their phonebooks. Our qualitative data, showing that many users planned to prune their lists after participating in the study, shows that there needs to be a way to control who gets added to the phonebook, or at least who is visible. Making it clearly visible which services are linked, how to unlink them, or providing controls to filter out specific contacts are all desired features.

Transient Contacts

An interesting finding was around the use of transient contacts: people who are only in your life for a very short amount of time. Perhaps this is someone you're buying a couch from off of Craigslist or the email of a teaching assistant for a class that will last 12 weeks. Having an easy way to add a contact for a set duration is important and will lessen the current "clutter" of large contact lists full of contacts that have not been contacted in years and were only relevant for a very short time.

Implications for Theory

Over the years, many theories relating to communication patterns have been created. In this section, we will discuss our findings with respect to other work that has addressed the size of social networks. For example, Dunbar's Number [8] states that humans can have approximately 100-200 stable social relationships. While we saw a mean of over 300 contacts, the median was 164, very close to this number. In later work, Hill and Dunbar [13] found that family represented 21% of those in one's social network. We found this percentage to be only 13%. This could be due to a larger number of "friends" and colleagues being imported through services such as Facebook or LinkedIn and can more accurately describe today's broader social connections where it is easier to minimally stay in touch with large numbers of more distant friends.

We can also look at theories related to close contacts, such as Marsden's Discussion Contacts [17]. In this work, he found that Americans typically have three contacts with which they discuss "important matters." We asked this question in our survey as well, and we found a higher number of "Discussion Contacts." For the median phonebook size, there were 10 contacts in the highest tie strength bucket, which included the question on discussing "important matters." Interestingly, despite there being more Discussion Contacts in our data, we found almost exactly the same percentage of family versus friends that were discussion contacts as compared to Marsden's research. Marsden found 1.5 kin and 1.4 non-kin (a ratio of 1.07) while we found a ratio of 1.05 despite having a higher percentage of friends in the phonebook as discussed above. Further research along these lines would be required to make strong claims here.

LIMITATIONS

While we believe that the scale of our analysis and the strong diversity of our participants enables us to make the strong quantitative claims above, there are a few limitations to our method.

First, only participants who had access to the web from a computer, and could install our application onto their smartphone could participate. However, as we saw an education distribution and urban/rural split that were almost identical to the US population, this does not seem to have affected recruitment.

Second, our study only contains participants from the United States. While many participants self-identified as immigrating from other countries, and we covered over 30 states with our participants, communication patterns in other parts of the world are likely different and would need to be studied independently.

Finally, this study only represents a snapshot of the phonebook as of Summer 2014. As mobile phonebooks change and as more services add integration to the phonebook and insert contacts from their services, the makeup of the phonebook will likely change. Also, the rise of services such as WhatsApp (which is still not extremely popular in America due to a lesser need to send international messages) likely will limit the usefulness of SMS logs for analysis over time and in other regions. Currently, the SMS app is the most used app of any category on Android in the US.

CONCLUSION

We have described a study investigating the composition of Android phonebooks and communication history with a variety of contacts. Through collecting data from 200 participants about their 65,940 contacts as well as metadata for 74,493 calls and 306,228 SMS messages we have

shown that mobile phonebooks are far more than a list of trusted contacts who are called and massaged frequently.

Since 29% of contacts are completely unknown and more than 80% of communication is to five people, this work has strong implications on the design of applications that use the phonebook to import contacts and for the design of mobile phonebooks that currently make it hard to access the most frequently contacted people in one's life.

Although Android is the dominant smartphone OS, it would be interesting to conduct this study on other phone platforms. Windows Phone has a very similar mechanism for applications to integrate with the phonebook and understanding differences in use for certain design differences between the platforms would be interesting. Overall, we expect the findings of our research to be useful for those designing new phonebooks as well as those creating new communications applications that take advantage of the contact list available on mobile platforms.

REFERENCES

1. Agathe Battestini, Vidya Setlur, and Timothy Sohn. 2010. A large scale study of text-messaging use. In Proceedings of MobileHCI 2010.
2. Frank R. Bentley, JoEllen Kames, Rafiq Ahmed, Rhiannon Sterling Zivin, and Lauren Schwendimann. 2010. Contacts 3.0: bringing together research and design teams to reinvent the phonebook. In CHI '10 Extended Abstracts.
3. Bentley, Frank, and Crysta J. Metcalf. "The Use of Mobile Social Presence." IEEE Pervasive Computing 8.4 (2009): 35-41.
4. Frank R. Bentley and Crysta J. Metcalf. 2007. Sharing motion information with close family and friends. In Proc. CHI '07. ACM, New York, NY, USA, 1361-1370.
5. Böhmer, Matthias, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. "Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage." In Proc. Mobile HCI '11.
6. Consolvo, Sunny, et al. "Location disclosure to social relations: why, when, & what people want to share." In Proc. CHI 2005.
7. Maeve Duggan and Aaron Smith Social Media Update 2013. 2013. http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/
8. R.I.M Dunbar. 1998. The Social Brain Hypothesis. Evo. Anthro., 6:178.
9. Eagle, Nathan. "Mobile phones as sensors for social research." Emergent technologies in social research. Oxford University Press, New York (2011): 492-521.
10. Furletti, Barbara, et al. 2014. "Use of mobile phone data to estimate mobility flows. Measuring urban population and inter-city mobility using big data in an integrated approach." In Proceedings of the 47th Meeting of the Italian Statistical Society.
11. Rebecca Grinter and Margery Eldridge. 2003. Wan2tlk?: everyday text messaging. In Proc. CHI 2003.
12. Harboe, Gunnar, et al. "Ambient social tv: drawing people into a shared experience." In Proc. CHI 2008.
13. Hill, Russell A., and Robin IM Dunbar. "Social network size in humans." Human nature 14.1 (2003): 53-72.
14. Ito, M., Mobile Phones, Japanese Youth, and the Re-Placement of Social Contact. In Proceedings of Annual Meeting for the Society for the Social Studies of Science, (Cambridge, MA, 2001).
15. Lisa Lambert. More Americans move to cities in past decade-Census. Reuters. Mar 26, 2012. http://www.reuters.com/article/2012/03/26/usa-cities-population-idUSL2E8EQ5AJ20120326
16. Ling, Rich, and Birgitte Yttri. "Nobody sits at home and waits for the telephone to ring: Micro and hyper-coordination through the use of the mobile telephone." Telenor Forskning og Utvikling, FoU Rapport 30 (1999): 99.
17. Marsden PV. 1987. Core discussion networks of Americans. Am. Sociol. Rev. 52:122-313
18. Natasha Lomas. 2014. Android Still Growing Market Share By Winning First Time Smartphone Users. Techcrunch. http://techcrunch.com/2014/05/06/android-still-growing-market-share-by-winning-first-time-smartphone-users/
19. Jun-Ki Min, Jason Wiese, Jason I. Hong, and John Zimmerman. 2013. Mining smartphone data to classify life-facets of social relationships. In Proc. CSCW '13. ACM, New York, NY, USA, 285-294.
20. Raento, Mika, Antti Oulasvirta, Renaud Petit, and Hannu Toivonen. "ContextPhone: A prototyping platform for context-aware mobile applications." Pervasive Computing, IEEE 4, no. 2 (2005): 51-59.
21. Catherine Rampell, Data Reveal a Rise in College Degrees Among Americans. New York Times. June 13, 2013. http://www.nytimes.com/2013/06/13/education/a-sharp-rise-in-americans-with-college-degrees.html
22. A. Schmidt, T, Stuhr, and H.W. Gellersen. 2001. Context-Phonebook - Extending Mobile Phone Applications with Context. In Proc. Mobile HCI 2001.
23. Jennifer VanGrove. 2012. Your address book is mine: Many iPhone apps take your data. Venture Beat. http://venturebeat.com/2012/02/14/iphone-address-book/
24. Jason Wiese, Jun-Ki Min, Jason I. Hong, and John Zimmerman. 2015. "You Never Call, You Never Write": Call and SMS Logs Do Not Always Indicate Tie Strength. In Proc. CSCW '15.