Building a Dynamic and Computational Understanding of Personal Social Networks

Jason Wiese, Jason I. Hong, John Zimmerman
Carnegie Mellon University
Pittsburgh, PA
{jwwiese, jasonh, johnz}@cs.cmu.edu

ABSTRACT

While individuals' personal social networks are extremely important in their day-to-day lives, computational systems lack meaningful representations of them. We argue that recent trends in computer-mediated communication, the ubiquity of smartphones, usage of online services, and new approaches to real-world social science experimentation have created an opportunity to dynamically generate representations of personal social networks that will be useful in a variety of application areas. We describe several preliminary steps we have taken to investigate this vision, which demonstrate that the approach appears to be feasible and seems likely to produce useful results. While there are significant privacy concerns in this space, we outline two approaches for dealing with them. Finally, we close with a discussion of several application areas that might benefit from this new process for generating representations of personal social networks.

Categories and Subject Descriptors

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

General Terms

Design, Human Factors

Keywords

Personal social network, smartphone, data mining, privacy

1. INTRODUCTION

It is difficult to overstate the importance of social relationships to the human species. Social relationships permeate every aspect of our personal and professional lives. Yet, the broad array of computing systems we interact with on a day-to-day basis have an extremely coarse representation of our personal social networks (PSNs), if any at all. If we could capture some computational representation of a particular user's PSN we could apply that knowledge to a broad range of application areas that might benefit from this information. Doctors might be able to use this information to diagnose and treat people with health conditions that are related to social interaction (such as clinical depression). Computing systems could leverage this information to include socially sensitive user interfaces, or to automate tasks such as specifying privacy and sharing settings. Users themselves can benefit from seeing representations of their own PSNs, to help them be more self-aware, and perhaps to help arouse behavioral change.

MCSS'12, June 25, 2012, Low Wood Bay, Lake District, UK.

However, our own PSNs are extremely difficult to quantify and describe and are constantly changing in ways that are sometimes subtle, and other times monumental. We may not even fully realize when our own PSNs are changing, and relationships develop and change as we go about our daily lives, regardless (and sometimes because) of where we are.

Despite the plethora of challenges in understanding the phenomena of interpersonal relationships, social scientists have striven to understand, describe, and quantify personal social networks, through various approaches mostly involving surveys and interviews that are often time-intensive both for the researcher and the participant. Questions can be extremely difficult to answer, and likely depend on numerous external factors. Truly, the effort expended to understand social relationships and PSNs is a testament to the importance of these concepts. Yet, even our most accurate representations of an individual's PSN fall short of the true PSN.

In today's exciting and rapidly changing technological landscape, we feel that there are four major trends that may enable computational systems to infer a "good enough" representation of a person's social graph:

1. increase in computer mediated communication
2. increasing ubiquity of smartphones
3. increase in use of online services that provide a rich history of actions
4. development of a new kind of social science that tests hypotheses in the real world, in real time

The smartphones people increasingly carry in their pockets, purses, and backpacks instrument the world [7], [12], [20][30] completely differently than the traditional embedded systems favored in UbiComp. These networked devices are packed with a variety of sensors including microphones, cameras, location (Wi-Fi, phone network, and GPS), movement (accelerometer and compass), and they most often move through the world generally attached to an individual. These devices can provide both individuals and service providers with rich data about where people are and what they are doing without the need to purchase any additional infrastructure.

Second, and not fully decoupled from the first trend is the continually increased usage of technology-mediated communication; that is, communication that replaces face-to-face interaction. People increasingly use a mobile phone as a communication hub, and this device generates and stores a detailed log of the who, when, and direction of the communication exchange. People also use email and social networking services, which also keep detailed logs of social actions including SMSs. These stores of metadata, both individually and in their totality, offer a great first start at inferring a dynamic personal social network via observation.

In addition to social interaction, people increasingly engage in many different types of online services that can provide details of the actions they are taking in the world. Shopping services like amazon.com and zappos.com provide insights on what people buy and for whom. Personal informatics services like Weight Watchers or MapMyRun provide insights on goals and actions taken in pursuit of these goals. Services like Tripit keep records of where people travel and who whom; online financial services keep details on how money comes in and how it get spent, and services such as Bing and Google keep track of search topics. Individually, these traces of actions provide only a sparse and brittle perspective of a person and their social cares and concerns. However, if aggregated together, they could provide a powerful new source of information for understanding the nature and the strength of the connections between people.

The increased number of actions taking place online has led to new approaches for testing new ideas in living, dynamic systems: A/B testing based on web analytics. Today many web service providers divide their users and present different variations of their applications in order to more systematically refine their services and to scientifically understand the relationship between their design choices and people's actions. As this type of real-time real-world social science matures, it seems natural that it will move beyond investigating behavior within a single service and will advance to include multiple services as well as behaviors that can be sensed on the smartphones so many people are carrying.

The combination of these trends, all of which are all enabled by or accelerated by the proliferation of mobile devices, presents a new opportunity to observe and model a person's personal social network with a level of detail never before imagined. While previous understandings of PSNs have relied on primarily qualitative self-report data, these new data sources provide exponentially more information, collected both automatically and in collaboration with users. These new, observed social networks will not longer need to be static representations. Instead, the observed models can dynamically change in reaction to changes people make in their own lives.

We have begun preliminary investigations to answer some of the most important questions that arise from this vision. As a first step, we collected self-report data on users' personal social networks, then determined the feasibility of inferring an abstracted representation of PSNs based on "felt closeness" from observable data. Further, we applied this information to the domain of sharing and privacy preferences to help us understand whether this abstraction of PSN is "good enough" for this application area. We have followed this up by collecting real behavioral data logs, which we are in the process of analyzing to build abstract representations of PSNs.

In the sections that follow, we describe various approaches to describing PSNs, and our vision for how we might build new and complementary representations. Next, we describe in further detail our preliminary investigations into the subject. We discuss privacy risks and issues associated with our data aggregation method, and how we might mitigate these risks. Finally, we close with an overview of some of the application areas and a discussion of why they are useful and how we might investigate them.

2. UNDERSTANDING PERSONAL SOCIAL NETWORKS

In order to infer a person's PSN, it is important to understand the basic concept of a PSN and how social scientists have previously generated representations. A person's PSN exists as a concept, not necessarily as a concrete representation. There is a lot that we do not know about PSNs: how they work, what their important dimensions are, what their limitations are. A broad range of researchers have made efforts to uncover the nature of PSNs and construct different representations of them. Our goal is to construct representations of PSNs that allow us to explore the areas above. We want to represent PSNs in ways that are functional, so our goal is not necessarily to generate the most accurate or complex representations possible.

The rest of this section contains a broad overview of different approaches to understanding PSNs and generating representations of them. Some representations have been constructed manually and statically, which tend to be both time intensive and holistic. Others are constructed automatically and use data that tend to be more siloed and less aggregated across sources. Finally, we discuss our vision, which involves some elements from both of these categories.

2.1 Manual/Static Representations of Social Networks

Here we describe approaches to constructing PSNs that involve intensive work by humans. These are often quite thorough, but static and difficult to recreate. The best known and most tested forms of understanding PSNs are intensive, in-depth tasks undertaken by researchers in social science, sociology, and anthropology. There have been numerous studies examining the frequency of various tie strengths. For example, a study of 3000 Americans showed that the average American has just four strong ties, most having between two and six [5]. Another study of about a thousand adults found that, on average, people have about 10 friends they meet or speak with at least weekly [5]. Spencer and Pahl conducted 60 "in-depth, face-to-face individual interviews" in an effort to understand individuals' personal communities [26]. They cover a broad range of issues including stages of relationships, general classes of relationships, and distinctions between chosen and given relationships.

A great deal of literature has demonstrated that the frequency of interaction between a pair increases their likelihood of forming a friendship or romantic relationship [2]. Some studies have employed physical proximity as a proxy for the amount of social interaction between pairs [14], [23], for example, showing that communication frequency drops exponentially with the distance between a pair [1], [31].

The literature on this subject is much more extensive than we could possibly cover here, but the resounding theme throughout is that the data collection process is manual and intensive, and generally relies on a combination of surveys and interviews.

2.2 Dynamic, Siloed Representations

In this section we discuss approaches that are based on computational data, and could therefore be automated. Recently, approaches both in the research and commercial sectors have taken to understanding personal social networks using the data available within a particular service. The main drawback to these approaches is the limitation of using a single or only a few data sources to represent a person's PSN.

In a series of papers, Eagle et al. [9], [10], [11] studied and modeled human social structure using mobility data from mobile phones. 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 did not always match mobile phone data, heavily depending on recency and salience of the interactions. They were also able to infer 95% of friendships among their participants based on call records and Bluetooth proximity. In other work, Choudhury and Basu [4] analyzed audio data using speech-processing techniques to infer social network structure.

There have also been early results in using communication data to model social graphs. For communication data, past work has developed techniques for inferring different groups from email and from online social network usage [6], [8], [13] [19], [22], [25] or inferring the strength of ties between individuals based on communication patterns [15], [29], and using mobile phone call data to model social structure [16].

These studies pave the way for understanding social relationships in a more dynamic, automated fashion. The reality mining work was an instrumental and pioneering piece of work in this domain. Additionally, the promising results by Gilbert and Karahalios [15] demonstrate the feasibility of unpacking complex dimensions of relationship using this kind of behavioral data. However, this work provides only a snapshot of a person's actual behavior, and generally does not combine information across many sources.

2.3 Multi-source Automated and Dynamic Representations

Our goal is to combine the dynamic approach just mentioned with aspects of the more classic holistic approach to produce more robust representations of PSNs in a lightweight and rapidly repeatable fashion. The realization of this vision depends on the automatic collection and synthesis of data from many varied sources, which can produce a dynamic model of the current state of a person's PSN that is reliable and continuously updated.

3. PRELIMINARY INVESTIGATIONS

Our recent and ongoing work has focused on exploring different aspects of this vision. Below we describe three steps of our work in this area. First, we discuss an initial feasibility study where we explored self-report data about the relationship between the to-be-sensed behavioral data, and different outcomes that the future system might predict. Next we describe a data collection that allowed us to use behavioral data to infer qualities of social relationships and predict sharing preferences. Finally, we discuss next steps for building models using the collected data.

3.1 Initial Feasibility Study

Collecting the data and building the system to realize this vision is a significant investment, and before we dove in, we wanted to understand what we might be able to tell about a person using this data that we could hypothetically collect. In an effort to understand the full picture of what we were proposing, we conducted a study online with 39 participants with three major components. While this study is fully described in [27], we describe the important components here.

First, participants provided us with a list of name of people who came to mind from different aspects of their lives, as well as a list of all of their Facebook friends. We took all of their friends from the first list and enough friends from the Facebook list to reach 70 total names per participant.

Next, participants answered a series of questions about each of their 70 friends. These questions targeted some information that we consider easy to obtain technologically (e.g., how old is this person, how often do you communicate with this person, etc.) as well as two questions in particular that were clearly not directly observable: "how close do you feel to this person?" and an open grouping task: "put this person into the group or group(s) in which they belong." Finally, participants went back though their list of 70 friends, this time specifying their willingness to share certain kinds of information with each of these 70 friends.

While all of the data we collected in this study was based on self-report, we were hoping to see that some of the answers we would not have been able to observe directly might be predicted by some of the data that we could observe. The results supported our hypothesis. Specifically, observable features, especially frequency of communication, were reasonably good predictors of both how close the participant felt to their friends, and how willing they were to share with those friends. Furthermore, closeness seems tied to sharing more so than the generated social groups across participants.

Another important finding that resulted from this work was the discovery of users' preferences for setting sharing preferences based on "in-common information." This means that a user's willingness to share her own information depends on a dynamic piece of context that relates to a particular person. For example, "I am willing to share my location with this person all of the time" was less favorable than "I am willing to share my location with this person when I am someplace that she has been before." This finding supports our hypothesis that the contextual data that we can collect from smartphones may be powerful for implementing solutions in the application areas we have identified.

3.2 Using Real Data to Map a Personal Social Network

Following from the previous study, our current goal is to understand how well these models might be built using real behavioral data. Our data collection process has several main components:

First, participants install an application on their existing Android smartphone which extracts data from the phone and stores it in a file that they can upload to the study website. This data includes phone call logs, sms/mms logs, and their complete contact list. In addition to uploading this data file to the website, participants also upload a ZIP file that they downloaded from Facebook containing names of all of their Facebook friends, as well as all of their wall posts and direct messages. Finally, for additional compensation participants are invited to upload their personal calendars from at least the last 6 months. All content is stripped out from these three sources, and what we are left with is a log of everyone they are connected with on Facebook and on their phone, and a log of social behaviors across several channels of communication.

Similar to the previous study, participants report people that are important and relevant to different aspects of their life, but additionally we are able to select friends for the list who are people that they have communicated with frequently across different media. Finally, just as last time, some friend names are randomly chosen from Facebook.

To maximize the value of the data, participants do have to explicitly link and resolve duplicates across their contact list and Facebook friend list, though in the long term we hope to automate that process.

Again, as in the last study, participants answer a list of questions about each of the 70 friends on their list, but this time the questions include several different ways of examining tie strength, and include fewer questions about sharing preferences.

We use the uploaded data to generate a broad variety of features that can be used as model inputs. The feature space includes measures of how frequently the participant communicates with each friend across different media, the average length of a conversation or message, the time of day when communication occurs, which of the communication channels a participant has used with a particular individual, and how much data the participant has for the particular contact entry of each friend.

Example machine learning features we are examining include:

- **Voice calls**: frequency of voice calls overall, frequency of voice calls to an individual, time of day and day of week of calls, location of voice calls, ratio of calls initiated to calls received, length of voice calls
- **SMS messages**: features similar to above (frequency of messages, time of day, etc), simple sentiment analysis, response rate to messages, response time
- **Facebook data**: person is a friend on Facebook, relative number of communications (wall posts or messages), estimates of years known, similarity in organizations, tagged photographs, total number of friends
- **Sex and age**: Similarity or difference in sex and age, how long participants have known each other

3.3 Building Useful Models

Following this data collection, the next step is to build a variety of different models to help answer some questions about the process itself. In particular we hope to answer some of these questions:

- Which dimensions of tie strength can we predict with these data? Can we predict some better than others?
- Which features are the most meaningful or predictive ones? Is it the combination of multiple features that make the models useful, or is there one feature in particular that is clearly most important?
- What is the added value of combining data sources? Do they tell us the same things about a PSN or is the combination truly valuable? How much of the picture does Facebook account for? Call logs? Contact list?
- In what situations do the models make errors? Are these systematic or seemingly random? Is there a good explanation for the cases where the model makes an incorrect prediction?
- Are the models themselves meaningful? Are they intuitive? Does the structure of the model reveal any interesting insights into the nature of PSNs?

Another important aspect of building these models is figuring out how particular types of data should be interpreted. First, is the straightforward idea that a piece of data (an instance of a phone call in the call log, for example) might directly reveal a piece of information about an interpersonal relationship (I called you on a Friday night, so you are somebody I enjoy social interaction with). However, these same kinds of data may also simply be pieces of context for understanding a different piece of data (for example, I made that phone call and received an SMS message from the same number shortly after, which may indicate something more interesting about the SMS message than if it was not preceded by the phone call).

4. HANDLING PRIVACY CONCERNS

However valuable a person's data is when it's siloed, it becomes much more valuable when streams of data can be linked together. It can be difficult for even the most qualified of experts to reason about the risks and benefits of aggregating data, let alone end-users, who are likely to be less comfortable with what the risks are and how to deal with them.

Privacy is a highly challenging area that is deeply embedded in legal, social, and technical norms. While we cannot guarantee that any of the techniques we explore will work or be accepted by end-users, it is essential to invest significant resources into working on this issue. One way we can improve the likelihood of finding a usable solution to the privacy issue is to take a multi-pronged strategy, where we investigate several approaches in early iterations of our work (where our studies will be done at small-scale), and adopt the best ones for later iterations.

One area to investigate is developing controls over and usable feedback on what data is gathered and shared with us. In recent work, Brush et al [3] conducted user studies to investigate the perceived utility of various kinds privacy-preserving techniques for location data, such as discretization, sampling, and Gaussian noise, and found that just removing data for a person's home was usually sufficient. Here, we will do similar work, expanding it for other kinds of sensed data. Other kinds of techniques we will also investigate include:

- Opting-out for capturing data at specific places or times.
- Letting people delete data before being shared with us.
- Letting people send samples rather than complete data sets
- Using place entropy as a way of helping people configure when sensed data should and should not be gathered (as has been done for location privacy preferences [27])

Another area to explore is to analyze the tradeoffs in doing local processing of data on people's smartphones. One approach in this area is local processing using only the end-user's own data, versus doing global processing with all of the data. For example, how effective and how accurate is the PSN using only one's personal data? The rationale here is that end-users may find this kind of approach more palatable, in that their data is never shared with anyone else.

Another approach is to do pre-processing of the data before it is shared with our centralized servers. For example, how effective will the models be if everything is first discretized by time and/or space? As another example, what if locations are turned into unique labels using one-way hash functions? Can we still infer things like entropy and co-locations while greatly minimizing where people are actually going?

We hope that by minimizing the amount of data that we need to collect to make the PSN useful, and maximizing user control for what data of theirs is collected, the benefits of having a computationally constructed PSN outweigh the privacy concerns.

5. APLICATION AREAS OF RICH PERSONAL SOCIAL NETWORKS

In this section, we describe several different application areas where we think the development of rich personal social networks have the potential to make a big impact.

5.1 Detecting Social Characteristics: Leadership and Depression

We have a strong hunch that leadership and depression are in fact detectable via social behavioral models and representations of

PSNs. One of the critical components of leadership can be found in communication. We expect that people who express this quality will exhibit different communication patterns across their social graph than others. Similarly, when looking for depression, we expect to observe a decrease in social activity across certain areas of a person's social graph, and we expect people to spend less time in highly social places outside of home and the workplace.

Also, leadership and depression are valuable qualities for organizations and for individuals to recognize. Detecting leadership qualities can help organizations develop more effective teams and management structures. In addition, it can work as a feedback mechanism to guide people to more effective behaviors. The ability to detect the onset of depression can help with early intervention. This can improve the quality of an individual's life and it can lead to more effective use of human resources within an organization. We suspect that for people undergoing depression, we will see people communicating with strong ties less often, less social interactions with others, and going out less often.

5.2 Supporting Management of Privacy and Sharing Policies

The lack of a detailed representation of PSNs prevents computational systems from effectively supporting social actions and leads to breakdowns such as accidental self-disclosure of information to the wrong group; the inability of others to access important information that an should have access to due to a person's security setting; and the ever increasing time and attention people must spend to engage with the individual sharing and security policies across the many different systems they interact with. To address this issue, we will develop a system that predicts willingness and desire to share specific kinds of information with specific individuals.

We will also examine the effectiveness of latent social role analysis in helping people to manage their privacy preferences. Using this technique, we can identify when a user's activity (as collected by mobile sensors) corresponds to her private social role and if personal information should not be disclosed. For example, for a friend finder application, a professor would like to share her location with her colleagues and students so that they can schedule meetings with her, know where to find her for discussions, or expect delays for scheduled meetings if she is still somewhere else on campus. However, when she is acting as a mother taking care of her children after work, such information should be kept private to her and her family and not to be disclosed to her colleagues.

5.3 Developing Concepts of Relationships and Personal Social Networks in Social Psychology

Currently, understanding and representation of interpersonal relationships and personal social networks in social psychology is limited by the resources of researchers in the field. While having access to these new models of PSNs will not replace traditional methods for understanding social phenomena, we do think that there is a distinct ability for them to complement each other, and this new tool might augment the process of studying PSNs.

For example the data that can be collected using the methods above can provide researchers with dynamic data that may indicate relationships that have changed for participants. People might be interacting more with some of their friends or less with others and not even realize it. Where a purely self-report approach might not surface these differences, augmenting the approach with this data could reveal such things. Similarly, this kind of new tool may reveal different dimensions or features of interpersonal relationships that have not yet been identified or closely studied. For example, if there are two people that I am close to and see and talk to every day, but one of them I also speak to on the phone quite often, is there something different about our relationship?

Furthermore, because of the burdensome nature of existing methods, data collections tend to be infrequent and spread out, even for studies that aim to understand relationships and how they change over time. With this new dynamic and automated method, perhaps we will be able to understand for the first time what relationships look like as they change. Do some people just stop calling, while others gradually decrease call frequency? Can you predict how a relationship will change? It is difficult to observe these behaviors with infrequent static methods, but with dynamic and automated methods perhaps we can learn more about the nature of human social behaviors.

5.4 Personal Informatics and Self Reflection

Personal informatics tools are tools that help people understand interesting aspects of their own behavior [20]. Using the data that we collect to construct representations of PSNs, there are several ways to present data to individuals.

One way is to take the raw streams of data and let the users view theses on their own. Example ideas here include estimating how much you walk every day, your commute, your inferred daily schedule, or how weather affects your travel patterns.

Another way to present the data is to show users representations of their own PSNs. Individuals can witness how their own relationships have grown or changed over time, and perhaps even compare their own PSN structure to other kinds of common structures. Are you a person who maintains lots of loose relationships, or a person who maintains a few strong relationships? Do your relationships with other people last for more time than the average person or less? An important aspect that we will investigate is letting people compare themselves to others, to see how normal they are (or abnormal, as the case may be). This also has the potential for helping people realize their vision for who they want to be (e.g.: the PSNs of leaders tend to look like X, so if I want to be more of a leader, maybe I should do things to make my PSN also look like X).

5.5 UI Personalization

If we can generate reliable and accurate representations PSNs, this information could be made accessible to application developers (subject to privacy concerns) for a variety of uses. For example, a user's feed aggregator might present shared news stories from her coworker that she just had dinner with more prominently than those of a coworker that she does not see often. Another example of this idea is that users could be notified of shared content in online social networks based on their relationship with that person in general, beyond the context of that specific social network.

One risk here is that users will be uncomfortable making this information accessible to developers who wish to improve their applications for fear of misuse. While privacy is likely a concern for users in this application, previous work suggests that a strong value proposition may help mitigate people's privacy concerns [18], which has been further supported in recent work [3].

6. CONCLUSION

We have presented an overview of our vision for building computational relationships of personal social networks for a variety of applications. Our approach addresses important limitations of previous approaches and has the potential for a broad impact on how individuals interact with their devices, how they interact with each other through their devices, and how we can use these representations for a variety of professional and research application areas. Based on current technological trends and our own work to examine the feasibility of this approach, we feel there is a powerful opportunity before us, with many interesting challenges and problems to address.

7. ACKNOWLEDGEMENTS

This work was sponsored by Defense Advanced Research Projects Agency Defense Sciences Office (DSO) Program: DCAPS HERO/N66001-12-C-4196, and supported in part by a Yahoo Key Scientific Challenge Award, a Google Research Award, and the National Science Foundation under Grant No. DGE-0903659.

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the Defense Advanced Research Projects Agency of the U.S. Government. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

8. REFERENCES

[1] Allen, T. 1966. Managing the flow of scientific and technological information.

[2] Bossard, J. 1932. Residential propinquity as a factor in marriage selection. *American Journal of Sociology* 38, no. 2: 219-224.

[3] Brush, A.J. Bernheim, John Krumm, and James Scott. 2010. Exploring end user preferences for location obfuscation, location-based services, and the value of location. In *Proc. UBICOMP 2010*, 95-104.

[4] Choudhury, Tanzeem, and Sumit Basu. 2004. Modeling Conversational Dynamics as a Mixed-Memory Markov Process. In *Advances in Neural Information Processing Systems 17*, ed. Lawrence K Saul, Yair Weiss, and Léon Bottou, 281-288. MIT Press.

[5] Christakis, Nicholas A., and James H. Fowler. 2009. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Little, Brown and Company.

[6] Danezis, George. 2009. Inferring privacy policies for social networking services. *Conference on Computer and Communications Security*: 5-10.

[7] Davidoff, Scott, Brian D Ziebart, John Zimmerman, and Anind K Dey. 2011. Learning Patterns of Pick-ups and Drop-offs to Support Busy Family Coordination. *In Proc. CHI '11*

[8] De Choudhury, Munmun, Winter A Mason, Jake M Hofman, and Duncan J Watts. 2010. Inferring relevant social networks from interpersonal communication. *In Proc WWW '10*

[9] Eagle, Nathan, and Alex Pentland. 2005. Social serendipity: Mobilizing social software. *IEEE Pervasive Computing* 4, no. 2: 28-34.

[10] Eagle, Nathan, and Alex Pentland. 2009. Eigenbehaviors: Identifying Structure in Routine. *Behavioral Ecology and Sociobiology* 63, no. 7: 1057-1066.

[11] Eagle, Nathan, Alex Pentland, and D Lazer. 2009. Inferring Social Network Structure using Mobile Phone Data. *(PNAS)* 106, no. 36: 15274-15278.

[12] Estrin, Deborah, David Culler, Kristofer S J Pister, and Gaurav Sukhatme. 2002. Connecting the Physical World with Pervasive Networks 1, no. 1: 59-69.

[13] Fang, Lujun, and Kristen LeFevre. 2010. Privacy wizards for social networking sites. *In Proc. WWW 2010*

[14] Festinger, L. 1950. Informal social communication. *PsychRev* 57, no. 5: 271-282.

[15] Gilbert, Eric, and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proc. CHI 09*

[16] Hidalgo, César A, and C. Rodriguez-Sickert. 2008. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications* 387, no. 12: 3017-3024.

[17] Hui, Pan, Richard Mortier, Michal Piorkowski, Tristan Henderson, and Jon Crowcroft. 2010. Planet-scale Human Mobility Measurement. In *ACM Mobisys HotPlanet*

[18] Iachello, Giovanni, and JI Hong. 2007. End-User Privacy in Human-Computer Interaction. *Foundations and Trends in Human-Computer Interaction* 1, no. 1.

[19] Jones, Simon, and Eamonn O'Neill. 2010. Feasibility of Structural Network Clustering for Group-Based Privacy Control in Social Networks. In *Proc. SOUPS 2010*.

[20] Li, Ian, AK Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In Proc. *CHI2010*, 557-566.

[21] Mitchell, Tom M. 2009. Computer science. Mining our reality. *Science* 326, no. 5960 (December): 1644-5.

[22] Neustaedter, Carman, A.J. Bernheim Brush, MA Smith, and Danyel Fisher. 2005. The Social Network and Relationship Finder: Social Sorting for Email Triage. In *Proc. CEAS 2005*

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

[24] Pentland, Alex. 2008. *Honest Signals: How They Shape Our World (Bradford Books)*. The MIT Press.

[25] Sarma, Anita, Larry Maccherone, Patrick Wagstrom, and James Herbsleb. 2009. Tesseract: Interactive visual exploration of socio-technical relationships in software development. *ICSE*: 23-33.

[26] Spencer, L., & Pahl, R. E. (2006). *Rethinking friendship: hidden solidarities today*. Princeton University Press.

[27] Toch, Eran, Justin Cranshaw, Paul Hankes Drielsma, Janice Y. Tsai, PG Kelley, James Springfield, LF Cranor, JI Hong, and Norman Sadeh. 2010. Empirical Models of Privacy in Location Sharing. In *Ubicomp 2010*.

[28] Wiese, Jason, P Kelley, L Cranor, Laura Dabbish, J Hong, and John Zimmerman. 2011. Are you close with me? Are you nearby? Investigating social groups, closeness, and willingness to share. In *Proc. Ubicomp 2011*.

[29] Xiang, R, Jennifer Neville, and Monica Rogati. 2010. Modeling relationship strength in online social networks. *In Proc. WWW 2010*

[30] Zimmerman, John, Anthony Tomasic, Charles Garrod, Daisy Yoo, Chaya Hiruncharoenvate, Rafae Aziz, Nikhil Ravi Thiruvengadam, Yun Huang, and Aaron Steinfeld. 2011. Field Trial of Tiramisu : Crowd-Sourcing Bus Arrival Times to Spur Co-Design. *In Proc. CHI 2011*

[31] Zipf, G. 1949. Human behavior and the principle of least effort. Cambridge, MA: Addison-Wesley.