Analyzing Facebook Privacy Settings: User Expectations vs. Reality

Yabing Liu
Northeastern University
Boston, MA, USA
ybliu@ccs.neu.edu

Krishna P. Gummadi
MPI-SWS
Saarbrücken/Kaiserslautern, Germany
gummadi@mpi-sws.org

Balachander Krishnamurthy
AT&T Labs–Research
Florham Park, NJ, USA
bala@research.att.com

Alan Mislove
Northeastern University
Boston, MA, USA
amislove@ccs.neu.edu

ABSTRACT

The sharing of personal data has emerged as a popular activity over online social networking sites like Facebook. As a result, the issue of online social network privacy has received significant attention in both the research literature and the mainstream media. Our overarching goal is to improve defaults and provide better tools for managing privacy, but we are limited by the fact that the full extent of the privacy problem remains unknown; there is little quantification of the incidence of incorrect privacy settings or the difficulty users face when managing their privacy.

In this paper, we focus on measuring the disparity between the desired and actual privacy settings, quantifying the magnitude of the problem of managing privacy. We deploy a survey, implemented as a Facebook application, to 200 Facebook users recruited via Amazon Mechanical Turk. We find that 36% of content remains shared with the default privacy settings. We also find that, overall, privacy settings match users' expectations only 37% of the time, and when incorrect, almost always expose content to more users than expected. Finally, we explore how our results have potential to assist users in selecting appropriate privacy settings by examining the user-created friend lists. We find that these have significant correlation with the social network, suggesting that information from the social network may be helpful in implementing new tools for managing privacy.

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation**]: Miscellaneous; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

General Terms

Measurement, Experimentation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
IMC'11, November 2–4, 2011, Berlin, Germany.
Copyright 2011 ACM 978-1-4503-1013-0/11/11 ...$10.00.

Keywords

Facebook, privacy, measurement, online social networks

1. INTRODUCTION

Nearly half of the users who have access to the Internet are members of some online social network network (OSN) [6], resulting in a fundamental shift in the patterns of context exchange over the Web. The result of this shift is that instead of just being content *consumers*, individual end users are now required to be content *creators* and *managers*. Today, for every single piece of content shared on sites like Facebook—every wall post, photo, status update, and video—the uploader must decide which of his friends, group members, and other Facebook users should be able to access the content. The per-user average of 130 friends and 80 groups and events [21]—compounded with the average 90 pieces of content uploaded per user per month [21]—has turned the task of simply managing access to content into a significant mental burden for many users. As a result, the issue of privacy on sites like Facebook has received significant attention in both the research community [12,25,27,29] and the mainstream media [1–5,8,9].

Our overarching goal is to improve the set of privacy controls and defaults, but we are limited by the fact that there has been no in-depth study of users' privacy settings on sites like Facebook. While significant privacy violations and mismatched user expectations are likely to exist, the extent to which such privacy violations occur has yet to be quantified. In this paper, we take the first steps towards addressing the problem by analyzing the current state of affairs on Facebook. In particular, we center our analysis around two questions:

- What are the *ideal* privacy settings desired by users? How close are these to the actual settings that users have?
- Is there potential to aid users in selecting the *correct* privacy settings for their content? Can we reduce the mental burden on users by automatically grouping others into meaningful groups for expressing privacy settings?

Since we wish to examine whether users' desired privacy settings differ from their existing settings, we need to ask users detailed questions, i.e., survey them. Thus, we design a survey (implemented as a Facebook application) that examines users' current privacy settings and queries users about their desired settings.[1] In order to scale to a significant number of Facebook users, we recruited users to participate in the study using Amazon Mechanical Turk (AMT). We automatically crawled the existing privacy settings of each piece of uploaded content for 200 users, resulting in 116,553 observations of existing privacy settings. Additionally, each of the users answered survey questions about their desired privacy settings for up to 10 of their photos, resulting in a total of 1,675 measurements of desired settings.

In brief, we find that the current Facebook privacy settings match users' expectations only 37% of the time, indicating that current settings are incorrect the majority of the time. More worrisome, when the settings are incorrect, they almost always tend to be *more open* than the users' desired settings, exposing the content to more users than expected. Additionally, we find that even when users have changed their default privacy settings, the modified settings only match expectations 39% of the time, indicating that even the users who are more privacy-aware have difficulty managing and maintaining their privacy settings correctly. Finally, we demonstrate how our results suggest a potential way forward by showing that many user-defined friend lists (similar to the Circles feature on Google+ [10]) have significant correlation with the structure of the social network. This suggests that the membership and maintenance of friend lists may be aided through the use of community detection algorithms [16, 19, 35, 36, 38].

The remainder of this paper is organized as follows: Section 2 provides a brief overview of related work on measuring privacy in OSNs. Section 3 describes our data collection methodology and data set statistics. Section 4 analyzes our survey data, focusing the relationship between the actual and desired privacy settings. Section 5 focuses on potential ways to aid user privacy management and Section 6 provides a concluding discussion.

2. BACKGROUND

In this section, we briefly provide background on Facebook's privacy model before discussing studies of OSNs privacy and studies that recruit users from AMT.

2.1 Facebook's privacy model

At the time we deployed the survey, Facebook allowed users to manage the privacy settings of uploaded content (photos, videos, statuses, links and notes) using five different granularities: Only Me, Specific People, Friends Only, Friends of Friends, and Everyone.[2] Specific People allows users to explicitly choose friends (or pre-created friend lists, discussed below) to share content with. The default or "recommended" privacy setting for all content is Everyone, meaning users share their content with all 750 million Facebook users [7] if they decline to modify their privacy settings.

Facebook allows users to re-use Specific People privacy settings via friend lists. Users create a friend list, add a subset of their friends to it, name it, and can then select the list as a basis for privacy control. Friend lists are private to the user who creates them, unless the user explicitly chooses to display them as part of his profile.

The granularity of privacy settings varies according to content type. Photos are grouped into albums, and privacy settings are specified on an album granularity (i.e., all photos in an album must have the same privacy setting). For the remaining content types, users can specify different privacy settings for each piece of content.

2.2 User privacy

Privacy is an emerging challenge in OSNs, and a number of researchers have examined different aspects of the privacy problem.

Researchers have examined the privacy model of existing OSNs, demonstrating that sites often leak numerous types of privacy information [12, 26, 29]. A number of papers report that users have trouble with existing extensive privacy controls, and are not utilizing them to customize their accessibility [28, 30, 40]. Other work surveys users' awareness, attitudes, and privacy concerns towards profile visibility and show that only a minority of users change the default privacy preferences on Facebook [11, 22]. However, they do not study to what extent the actual selected settings match users' preferences. There is also significant work that explores new approaches that can enhance the content sharing privacy on OSNs [14, 15, 18, 20, 39].

There are several closely related papers which have measured the privacy considerations of different kinds of information, such as News Feed [23], tagged photos [13], basic profile information [32]. All of these papers demonstrate the importance of the ease of information access in alleviating users' privacy concerns. Madejski et al. [32] show that privacy settings for uploaded content are often incorrect, failing to match users' expectations. There are two primary distinctions between their work and ours. First, they rely on text analysis to select content that is potentially privacy-sensitive; doing so locates additional privacy violations but prevents an overall estimate of the fraction of content that has incorrect settings. Second, we directly compare the user survey results to the in-use settings, instead of relying on inferring the existing privacy setting though fake accounts.

2.3 Using Amazon Mechanical Turk

Most prior work uses small-scale surveys of locally recruited users to study user attitudes towards privacy. This approach affords more control over the surveyed population, but also limits the scalability of the survey. In our work, we take a different approach, recruiting users from Amazon Mechanical Turk (AMT), which offers the potential of greater scalability and a lower cost of running experiments [33]. We now give a brief overview of other studies that have recruited users from AMT.

There have been multiple studies showing that the behavior of participants on AMT is comparable to the behavior of laboratory subjects in more traditional economic and psychological experiments [24, 37]. Considering that compensation may affect the quality of survey results, Mason and Watts [34] show that in online peer production systems like AMT, increased financial incentives increase the quantity, but not the quality of work performed by participants. These studies provide evidence that AMT offers a

[1] This study was conducted under Northeastern University Institutional Review Board protocol #10-10-04.

[2] Facebook has since simplified their privacy setting options, presenting only Friends and Everyone by default. The other options are still available via Custom settings.

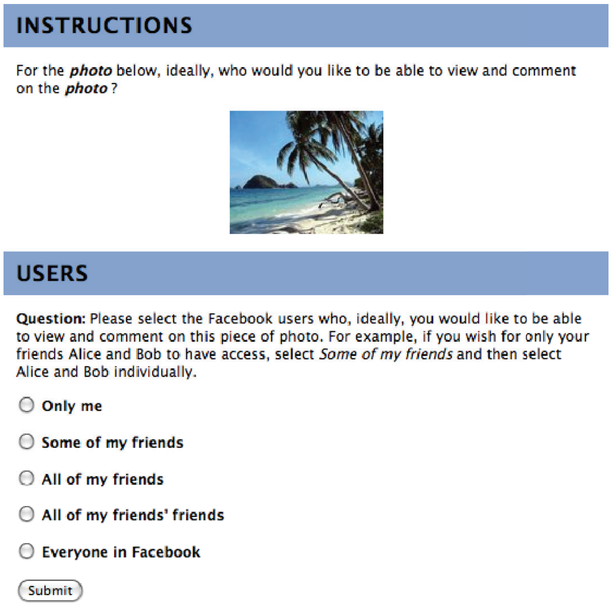


Figure 1: Screenshot of our survey. Each user was asked about 10 different uploaded photos.

potentially attractive way of quickly recruiting significant numbers of survey users.

3. METHODOLOGY

We now describe our approach for collecting data from Facebook users concerning privacy settings. We then detail a few statistics of the collected data set, and examine the demographics of the users who participated in our survey.

3.1 Approach

Our survey was hosted on a web server located at Northeastern University, and is available at `http://socialnetworks.ccs.neu.edu/yabing`. We designed our survey as a Facebook application. By doing so, the application is able to query Facebook to select content to query the user about, as well as to collect the current privacy settings for the user's uploaded content. It is important to note that all data collected is immediately hashed and anonymized; no non-anonymized data is ever written to disk.

When the user begins the survey, he is shown a consent form detailing the purpose and methodology of the experiment and asked to provide optional demographic information (age, gender, income, education level, and U.S. state). Then, the user is asked to answer questions about the ideal privacy settings of some of his uploaded content. Finally, the survey collects information from the user's profile, including the privacy settings for all uploaded content (photos, videos, statuses, links, and notes), any user-created friend lists, and the structure of the user's one-hop social network (i.e., the friendship connections between the user's friends).

The survey selects 10 photos to query the user about. In order to ask the user about both benign and potentially privacy-sensitive photos, the survey first randomly selects up to 5 photos that have non-default privacy settings (i.e., photos where the user has previously modified the privacy settings). Then, the survey chooses the remaining photos randomly from among all photos uploaded, regardless of privacy settings. For each photo, the survey asks the user who, ideally, should be able to view and comment on the photo. The user is presented with a number of options, which approximate the privacy settings currently allowed by Facebook (abbreviations are used in the remainder of the paper):

- **Only Me (Me)** Indicating that the photo should be private the user.
- **Some Friends (SF)** The user is asked which of his friends should be able to access the photo. The user can select friends individually from a list, or can specify users using any friend lists they have created.
- **All Friends (AF)** Indicating that all of the user's friends should be able to access the photo.
- **Friends of Friends (FoF)** Indicating that all of the user's friends, and all of their friends, should be able to access the photo.
- **Everyone (All)** Indicating that all Facebook users should be able to access the photo.

A screenshot of our survey is shown in Figure 1.

One of the benefits of building the survey as a Facebook application is the ability to quickly attract a large number of diverse users. To do so, we recruited users using AMT. We posted a Human Intelligence Task (HIT) describing the application, and offered users $1 to add our application and complete our survey. On average, we found that our survey took users 6 minutes and 30 seconds to complete, implying that completing our survey represented an average hourly wage of $9.23.[3]

There are a few limitations of our methodology that are worth addressing. We focus on photos, as these represent the most commonly uploaded content on Facebook and they have the most diverse privacy settings. Additionally, we only focus on content that is uploaded by the surveyed user; content uploaded by other users (even if it concerns the surveyed user) is not considered. However, Facebook applications are able to access all content uploaded by the user (i.e., the user cannot have a set of more privacy-sensitive photos that are hidden from applications). Finally, we treat each photo equally (in terms of the impact of an incorrect privacy setting), even though certain photos are likely to be more privacy-sensitive than others.

3.2 Data statistics

We now provide a brief overview of the data set that was collected. We deployed our HIT to AMT on May 2nd, 2011, and 200 users completed the survey. These users had an average of 248 friends, and had uploaded an average of 363 photos, 185 status updates, 66 links, 3 notes, and 2 videos. Only 45 users had uploaded fewer than 10 photos (of which 7 users had uploaded none). 81 out of our 200 users had also created at least one friend list, with a total of 233 observed friend lists. Thus, the average user who had created at least one friend list had 3 friend lists.

3.3 User demographics

One potential concern with recruiting users from AMT is the issue of bias; these users are unlikely to be a random

[3] We chose the compensation rate to be in line with recommendations from existing literature, which recommends paying close to the U.S. minimum wage of $7.25 per hour [33].

Type	Count	Me	SF	AF	FoF	Net	All
Photo	65,182	<0.1%	17%	37%	18%	1.3%	26%
Video	428	0.5	5.6	32	11	3.5	48
Status	37,144	0.1	9.7	35	4.5	3.4	47
Link	13,197	<0.1	5.4	36	9.2	2.0	47
Note	602	0.5	6.3	28	5.8	9.8	50
Total	116,553	<0.1%	13%	36%	13%	2.0%	36%

Table 1: Existing privacy settings for all content items. The different content types possess similar privacy setting distributions, and the default (All Facebook users) is selected for the plurality of the items.

sample of the Facebook population. We first note that the issue of bias is fundamental in user studies (e.g., psychology studies often use college students, another biased population), and our survey is no different. Nevertheless, we use the self-reported demographics in order to help to understand the nature and distribution of our user population.

In total, 195 (98%) users answered all demographic questions. We restricted our AMT user population to users only in the United States, and we had users from 40 of the 50 U.S. states. The most popular states were California (11%), Florida (11%), and New York (8.2%). We observed a slight male bias, with 54% of our users self-reporting as male; this differs from the overall U.S. Facebook breakdown of 42% male [17]. The self-reported age ranged between 18 and 60, with the median age being 24; this distribution is in-line with the overall U.S. Facebook population [17]. Finally, the self-reported yearly income level ranged from \$0 to more than \$120,000, with the median being \$10,000–\$20,000. These results demonstrate a wide variety of users, and are consistent with prior studies of AMT users [34].

One additional concern with our recruitment methodology is that our AMT users might be a "close-knit" group of friends, and not a more general sample of the user population. To evaluate whether this is the case, we examine how closely related our users are by examining the number of users who are friends on Facebook, and the number of user pairs with at least one common friend. Out of the 19,900 pairs of users $[\binom{200}{2}]$, 11 (0.05%) were direct friends and 13 (0.07%) were not direct friends but had at least one friend in common. Thus, our user population is not biased towards one small region of the Facebook social network.

4. ANALYSIS OF PRIVACY SETTINGS

In this section, we begin by investigating the distribution of user-selected privacy settings. We then use our user survey to compare the *desired* privacy settings and *actual* privacy settings, quantifying the frequency of discrepancies. Finally, in the following section, we examine the potential for aiding users in managing privacy by automatically grouping related users.

4.1 Existing privacy settings

We begin our analysis by examining the distribution of existing privacy settings. For each user who completed our survey, we collected the current privacy settings for all of their uploaded content (photos[4], videos, statuses, links, and notes). Table 1 presents an overview of the aggregated privacy settings.[5]

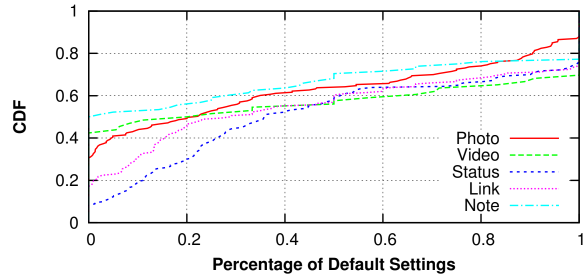


Figure 2: Cumulative distribution of the fraction of each user's content that remains at the default privacy setting, for the five different content types. The distribution differs across the content types; with many users only having changed the settings for a subset of their content.

We make two observations. First, out of 116,553 content items, 41,437 (36%) are shared with default privacy settings, meaning they are each visible to over 750 million Facebook users. This fraction is significantly higher than users indicated they desire (20%, discussed below in Table 2), suggesting that the users have not bothered to change the privacy setting from the default. Second, while the various content types show similar distributions, we note that photos have the most privacy-conscious setting: photos have the highest fraction of Some Friends, All Friends, and Friends-of-Friends, and the lowest fraction of Everyone. This suggests that users are more aware of the privacy settings for photos, implying that our survey below (which focuses exclusively on photos) is likely to underestimate the frequency of privacy violations for other types of content (since we observed that other types of content are much more likely to have default settings).

Next, we take a closer look at the per-user settings distribution in order to determine the fraction of users who have changed none, some, or all of their privacy settings from the default. To do so, we calculate the fraction of each user's content that remains at the default setting; Figure 2 presents the cumulative distribution of this fraction across our 200 users. We observe that the fraction of users who have changed either all or none of their privacy settings varies according to content type: for photos, this fraction is 43%, while for notes this fraction is 74% (implying that for photos, for example, 57% of people have some, but not all, of their photos shared with the default privacy setting).

4.2 Desired privacy settings

We now turn to examine the privacy settings that are desired by users, with a focus on comparing the desired settings with the current privacy settings. Recall that to measure the users' desired settings, we survey users concerning up to 10

[4]We do not consider the special album "Profile Pictures", as the user's profile picture is required to be publicly visible. We also disregard a total of 60 additional photo albums containing 7,540 photos for which the Facebook API returned uninterpretable privacy settings.

[5]Note that we also include the legacy **Networks (Net)** setting, indicating that all users in the same network (e.g., university or workplace) should be able to access the photo. This setting can no longer be selected by all users.

Actual setting	Desired setting Me	SF	AF	FoF	All	Total
Me	3	5	2	3	2	15
SF	3	12	28	3	0	46
AF	38	2	184	25	42	291
FoF	16	8	80	15	22	141
All	46	23	171	56	118	414
Total	106	50	465	102	184	907

Table 2: Comparison of the actual privacy settings and desired privacy settings for randomly-selected photos. The shaded cells represent instances where two are the same; this only occurs in 332 (37%, ± 3.14%) photos. When the two are different, they are more often shared with more users than desired (443 photos) than fewer users than desired (132 photos).

Actual setting	Desired setting Me	SF	AF	FoF	All	Total
Me	2	6	4	0	4	16
SF	2	12	29	8	11	62
AF	40	8	237	40	69	394
FoF	39	17	148	45	47	296
All	0	0	0	0	0	0
Total	83	43	418	93	131	768

Table 3: Comparison of the actual privacy settings and desired privacy settings for photos with non-default privacy settings. The shaded cells represent instances where two are the same; this only occurs in 296 (39%, ± 3.45%) photos. When the two are different, they are shared with more users than desired (254 photos) with approximately the same frequency as fewer users than desired (218 photos).

of their uploaded photos. Of our 200 surveyed users, 193 (97%) had at least one photo (and could therefore answer at least one survey question) and 155 (78%) had at least 10 photos (and could therefore answer all 10 survey questions). Additionally, Facebook also offers users the option of sharing photos with Networks [41]. We disregard this feature because many users are not members of networks and are unable to select this setting; this affects approximately 1.3% of photos. In total, our users answered questions concerning 1,675 photos (907 randomly selected photos and 768 random photos with non-default privacy settings).

It is important to note that while we selected photos independent of the albums to which they are assigned, Facebook's privacy settings are *per-photo album* rather than *per-photo*. We now briefly examine how many albums our random photo selection strategy covered. In total, the randomly selected photos came from 578 distinct albums. Our users had a total of 752 albums, meaning that we covered over 76% of all possible albums. Similarly, the non-default-privacy-setting photos came from 449 distinct albums out of 586 total non-default-privacy-setting albums, for a similar coverage of over 76% of all possible albums. Thus, our strategy of randomly selecting photos did not bias our survey towards a minority of the albums.

We divide our analysis into two parts, first focusing on a random selection of photos, and then focusing on photos with non-default privacy settings.

4.2.1 Randomly-selected photos

Table 2 presents the results of our survey for the 907 randomly selected photos, counting the number of photos with each combination of desired setting (columns) and actual setting (rows). First, we observe that for only 332 (37%, ± 3.14%[6]) of photos do the actual and desired settings match; indicating that 63% of the time, current privacy settings do not match users' expectations. Second, we observe that if we focus on the 575 photos that have incorrect privacy settings, 443 (77%, ± 3.44%) of them are shared with more users than desired. Third, and most worrisome, 296 (51%, ± 4.09%) of the 575 photos with incorrect privacy settings are incorrectly shared with all 750 million Facebook users. Taken together, our observations indicate that the problem of privacy management is endemic on Facebook—nearly two out of three of photos have incorrect privacy settings, and over half of these are incorrectly shared will all other Facebook users.

[6]All reported confidence intervals represent 95% confidence intervals.

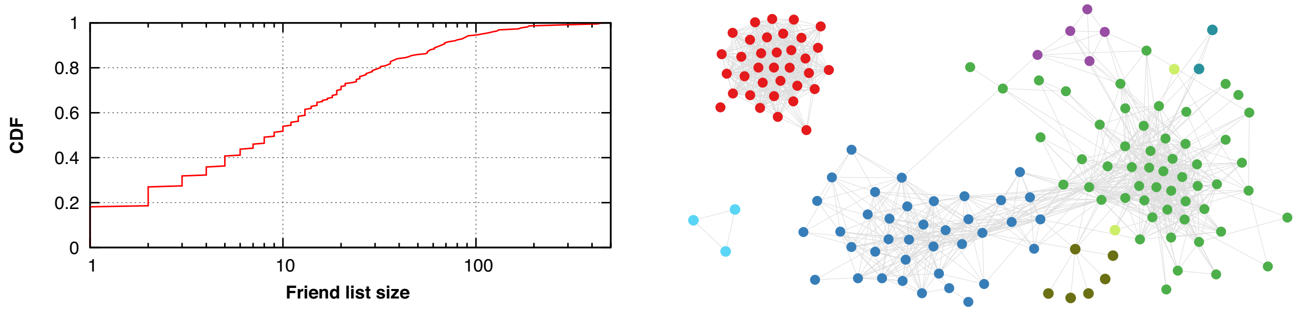
4.2.2 Photos with non-default privacy settings

One cause of the observations in the previous section are poor defaults: since it is known that users do not always adjust default settings, many of the photos could have incorrect settings because users have not bothered to adjust them. In order to shed light on the frequency of default settings causing privacy violations, we turn to examine only those photos which have non-default privacy settings. Since these photos, by definition, have had their privacy settings adjusted, we can see if the adjusted privacy settings better match users' expectations.

Table 3 presents the survey results for the 768 photos with non-default privacy settings. We observe that the fraction of correct privacy settings (296 photos or 39%, ± 3.45%) is approximately the same as the randomly selected photos. This indicates that even photos where the user has explicitly adjusted privacy settings still do not match the users' expectations the majority of the time. However, we also observe that the fraction of incorrect photos that are shared with more users than expected (54%, ± 4.50%) is much more even, when compared to the same fraction for randomly selected photos (77%, ± 3.44%). This suggests that while poor privacy defaults cause photos to be shared with more users than expected, users who are cognizant enough to modify their settings still have significant difficulty ensuring their privacy settings match their expectations.

4.2.3 Summary

Our analysis reveals that while users are uploading significant amounts of content to Facebook, almost half of the content is shared with the default privacy settings, which expose the content to all Facebook users. Users in our survey reported that this was the desired setting only 20% of the time, suggesting that the default settings are poorly chosen. More worryingly, even for photos for which the privacy settings have been modified by the user, the modified privacy settings match users expectations less than 40% of the time. This strongly suggests that users are having trouble correctly configuring their privacy settings and calls for new tools to manage privacy.



(a) (b)

Figure 3: (a) Cumulative distribution of the sizes of observed friend lists and (b) the 8 automatically detected groups of friends from one of the authors' Facebook social network. Nodes represent the author's friends and links exist between pairs of friends who are also friends. Nodes with the same color are automatically grouped together by the community detection algorithm.

5. IMPROVING PRIVACY TOOLS

As our final point of analysis, we examine the potential for assisting users in managing their privacy. Specifically, we focus on friend lists, a mechanism for users to group their friends that is similar to the Circles feature of Google+. We explore whether the friend lists could be automatically populated using community detection algorithms [16,19,35, 36,38] over the social network.

To do so, we examine the friend lists of our 200 surveyed users using the Facebook API. The cumulative distribution of the sizes of the 233 friend lists we examine is shown in Figure 3(a). More than 50% of friend lists have more than 10 members, while 20% of the lists have more than 30 members, which indicates the potential difficulties with manually generating and maintaining such large lists of friends.

One potential solution to the challenge of privacy management lies in leveraging the social links between the friends of a user to automatically group them into communities, where each community of friends can be used to create a friend list. We illustrate this in Figure 3(b), where we used a community detection algorithm [16] to automatically group the 144 Facebook friends of one of the authors into 8 friend lists.

For this approach to work effectively, users' friend lists need to correspond to tightly-knit communities in the network graph. To verify the extent to which users in friend lists form closely connected communities, we examine the normalized conductance [31] of the existing friend lists, whose value ranges from -1 to 1, with strongly positive values indicating significant community structure. Prior studies of social network graphs have found that normalized conductance values greater than 0.2 correspond to strong communities, that could be detected fairly accurately by community detection algorithms [31]. We analyzed the conductance values for our 233 friend lists and we found a significant positive bias. Over 48% of the friends lists have values larger than 0.2, suggesting that a large fraction of friend lists could be automatically inferred from the social network.

6. CONCLUDING DISCUSSION

We now briefly discuss a few issues brought up in the preceding analysis.

Automatically updating friend lists The results in Section 5 suggest that the social network can be automatically leveraged to aid users in selecting groups of friends to share content with. In ongoing work, we are developing Facebook applications that use the social network to help users generate friend lists conveniently. This is complementary to recent work on privacy "wizards" [18], which uses machine learning algorithms to infer communities. One potential advantage of leveraging the structure of the social network is the potential to easily update the friend lists as the user forms or breaks friendships.

Measuring privacy In general, privacy is a hard thing to measure, especially since it's hard even for users themselves to quantify. For example, photos alone are likely to have wildly varying privacy requirements, depending on who is in the photo, where it was taken, etc. In our survey, we simply treated all privacy violations as being equal, even though this is certainly not true in practice. In future work, we will explore mechanisms for measuring the "importance" of the various privacy violations, potentially by asking the users or using machine learning approaches on content metadata.

Additionally, when measuring the users' ideal privacy settings, we are treating the users' answers as ground truth. This may not always be the absolute ground truth, as the users' answers may vary with time (as social relationships change), or the users' may have not fully thought through the implications of a given setting. However, other user studies [18] are subject to the same limitation.

Reasons for incorrect settings Due to space constraints, we refrain from exploring *why* the privacy settings were incorrect. However, we note that such a study is non-trivial: just a few of the reasons for privacy violations include poor human–computer interaction mechanisms, the static nature of privacy settings, and the significant amount of work forced on the user to maintain the privacy of their content. We leave a full exploration of these to future work.

Acknowledgements

We thank the anonymous reviewers and our shepherd, David Wetherall, for their helpful comments. This research was supported in part by NSF grants IIS-0964465 and CNS-1054233, and by an Amazon Web Services in Education grant.

7. REFERENCES

[1] Facebook Stirs Privacy Concerns Again. *The New York Times*, 2010. http://gadgetwise.blogs.nytimes.com/2010/04/27/facebook-stirs-privacy-concerns-again.

[2] Facebook Executive Answers Reader Questions. *The New York Times*, 2010. http://bits.blogs.nytimes.com/2010/05/11/facebook-executive-answers-reader-questions/.

[3] Is There Life After Facebook? *The New York Times*, 2010. http://bits.blogs.nytimes.com/2010/05/12/is-there-life-after-facebook/.

[4] Price of Facebook Privacy? Start Clicking. *The New York Times*, 2010. http://www.nytimes.com/2010/05/13/technology/personaltech/13basics.html.

[5] Marketers Can Glean Private Data on Facebook. *The New York Times*, 2010. http://www.nytimes.com/2010/10/23/technology/23facebook.html.

[6] Global Publics Embrace Social Networking. *PewResearchCenter*, 2010. http://pewglobal.org/2010/12/15/global-publics-embrace-social-networking/.

[7] Facebook Climbs Toward 700 Million Users Worldwide, with Steady Growth in the US. *Inside Facebook*, 2011. http://www.insidefacebook.com/2011/04/21/facebook-climbs-toward-700-million-users-worldwide-with-steady-growth-in-the-us/.

[8] Facebook Facelifts Its Privacy Policy. *The New York Times*, 2011. http://gadgetwise.blogs.nytimes.com/2011/03/01/facebook-facelifts-its-privacy-policy/.

[9] Social Networks Offer a Way to Narrow the Field of Friends. *The New York Times*, 2011. http://www.nytimes.com/2011/05/10/technology/10social.html.

[10] Organizing My Online Friends. *The New York Times*, 2011. http://bits.blogs.nytimes.com/2011/07/12/organizing-my-online-friends/.

[11] A. Acquisti and R. Gross. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. *PET*, 2006.

[12] S. Ahern, D. Eckles, N. Good, S. King, M. Naaman, and R. Nair. Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing. *CHI*, 2007.

[13] A. Besmer and H. R. Lipford. Moving Beyond Untagging: Photo Privacy in a Tagged World. *CHI*, 2010.

[14] L. Banks and S. F. Wu. All Friends Are Not Created Equal: An Interaction Intensity Based Approach to Privacy in Online Social Networks. *CSE*, 2009.

[15] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin. Persona: an online social network with user-defined privacy. *SIGCOMM*, 2009.

[16] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebre. Fast unfolding of community hierarchies in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10, 2008.

[17] P. Corbett. Facebook Demographics and Statistics Report 2010. 2010. http://www.istrategylabs.com/2010/01/facebook-demographics-and-statistics-report-2010-145-growth-in-1-year.

[18] L. Fang and K. LeFevre. Privacy Wizards for Social Networking Sites. *WWW*, 2010.

[19] S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486, 2010.

[20] P. W. L. Fong, M. Anwar, and Z. Zhao. A privacy preservation model for Facebook-style social network systems. *ESORICS*, 2009.

[21] Facebook Statistics. http://www.facebook.com/press/info.php?statistics.

[22] R. Gross and A. Acquisti. Information Revelation and Privacy in Online Social Networks (The Facebook case). *WPES*, 2005.

[23] C. M. Hoadley, H. Xu, J. J. Lee, and M. B. Rosson. Privacy as Information Access and Illusory Control: The Case of the Facebook News Feed Privacy Outcry. *Electronic Commerce Research and Applications*, 9(1), 2009.

[24] J. J. Horton, D. G. Rand, and R. J. Zeckhauser. The Online Laboratory: Conducting Experiments in a Real Labor Market. *National Bureau of Economic Research*, w15961, 2010.

[25] B. Krishnamurthy and C. E. Wills. Characterizing Privacy in Online Social Networks. *WOSN*, 2008.

[26] B. Krishnamurthy and C. E. Wills. On the Leakage of Personally Identifiable Information Via Online Social Networks. *WOSN*, 2009.

[27] B. Krishnamurthy. I know what you will do next summer. *CCR*, 40(5), 2010.

[28] C. Lampe, N. B. Ellison, and C. Steinfield. Changes in Use and Perception of Facebook. *CSCW*, 2008.

[29] K. Lewis, J. Kaufman, and N. Christakis. The Taste for Privacy: An analysis of College Student Privacy Settings in an Online Social Network. *Computer-Mediated Communication*, 14(1), 2008.

[30] H. R. Lipford, A. Besmer, and J. Watson. Understanding Privacy Settings in Facebook with an Audience View. *UPSEC*, 2008.

[31] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. You are who you know: Inferring user profiles in Online Social Networks. *WSDM*, 2010.

[32] M. Madejski, M. Johnson, and S. M. Bellovin. The Failure of Online Social Network Privacy Settings. Department of Computer Science, Columbia University, Technical Report CUCS-010-11, 2011.

[33] W. Mason and S. Suri. Conducting Behavioral Research on Amazon's Mechanical Turk. *SSRN*, 2010.

[34] W. Mason and D. J. Watts. Financial Incentives and the Performance of Crowds. *KDD*, 2009.

[35] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69(6), 2004.

[36] M. E. J. Newman. Modularity and community structure in networks. *PNAS: Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 2006.

[37] G. Paolacci, J. Chandler, and P. G. Ipeirotis. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 2010.

[38] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *PNAS: Proceedings of the National Academy of Sciences of the United States of America*, 101(9), 2004.

[39] F. Stutzman and J. Kramer-Duffield. Friends Only: Examining a Privacy-Enhancing Behavior in Facebook. *CHI*, 2010.

[40] K. Strater and H. R. Lipford. Strategies and Struggles with Privacy in an online Social Networking Community. *BCS-HCI*, 2008.

[41] C. Wilson, B. Boe, A. Sala, K. P.N. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. *EuroSys*, 2009.