The Failure of Online Social Network Privacy Settings*

Michelle Madejski[†]
mgm2129@columbia.edu

Maritza Johnson
maritzaj@cs.columbia.edu

Steven M. Bellovin
smb@cs.columbia.edu

CUCS-010-11

Abstract

Increasingly, people are sharing sensitive personal information via online social networks (OSN). While such networks do permit users to control what they share with whom, access control policies are notoriously difficult to configure correctly; this raises the question of whether OSN users' privacy settings match their sharing intentions. We present the results of an empirical evaluation that measures privacy attitudes and intentions and compares these against the privacy settings on Facebook. Our results indicate a serious mismatch: every one of the 65 participants in our study confirmed that at least one of the identified violations was in fact a sharing violation. In other words, OSN users' privacy settings are incorrect. Furthermore, a majority of users cannot or will not fix such errors. We conclude that the current approach to privacy settings is fundamentally flawed and cannot be fixed; a fundamentally different approach is needed. We present recommendations to ameliorate the current problems, as well as provide suggestions for future research.

1 Related Work

To the best of our knowledge this is the first attempt to measure the correctness of privacy settings by first surveying user's sharing intentions, to aid the process of identifying potential violations, then confirming the potential violations with the user. We argue that this method produces a more accurate evaluation compared to passive data collection. This work draws upon many themes including: research on online social network (OSN) usage, surveys on privacy attitude, and evaluations of users' ability to manage access control policies.

A closely related study utilized survey analysis and data mining to study users' privacy attitudes, Facebook usage, and knowledge of profile visibility [1]. 209 Facebook users were surveyed on their knowledge of the visibility of their profile to strangers and network members, and on whether their profile was searchable. Mined profile data was used as ground truth for the questions. 8% of participants revealed more than they reported, 11% revealed less than they reported, and 77% revealed exactly what they reported. It is important to note that Facebook's features have changed dramatically since 2006 with the introduction of third party applications, the newsfeed, photos, videos, status updates, notes, and the ability to tag other users in most posted information. This study was a follow-up to an earlier study on the amount of information shared on Facebook [4]. The analysis was conducted by downloading 4,540 Facebook profiles and revealed that the majority of users shared a large amount of personal information. Very few users chose to restrict other users' ability to search for their profile (1.2%). Even fewer users chose to limit access to their profile to just friends (0.06%).

Joinson *et al.* researched users' motivations for using Facebook [6]. In a study of 241 Facebook users the majority of participants reported they utilize Facebook for "keeping in touch" with people with whom they have an offline relationship with, this includes looking up information about friends and communicating with

[*] This work was supported by NSF Grant CNS-0905403.

[†] Work done while at Columbia University.

friends. They found that users' privacy settings varied based on their reason for using Facebook. This point is critical to our evaluation – OSNs serve a purpose for users, which includes facilitating sharing information. An investigation of privacy settings is incomplete without understanding what users want to share.

Besmer and Lipford investigated users' attitudes towards photo sharing on Facebook, specifically when the photos are incidental data uploaded by a friend and the friend chose to tag the user [3]. The user's options are to untag the photo or leave the tag (allowing the association of the photo with the user's profile). Feedback from focus groups indicate that users have conflicting feelings about whether the owner of the photo or the people tagged in the photo should have the ability to dictate privacy settings. Currently, the photo owner is in full control of the photo. Thomas *et al.* suggested a mechanism that takes into account the privacy settings of all parties in these instances [15].

1.1 Privacy Surveys

Beginning in 1974, Westin conducted over 30 privacy surveys to measure privacy attitudes. Three general categories were used: fundamentalist (high concern), pragmatist (medium concern), and unconcerned (low) [9]. The majority of the participants were either fundamentalists or pragmatists, each year only a small portion of the participants were unconcerned. The most important factor was typically reported to be the ability to control one's own data.

The Westin surveys were conducted before OSNs were popular. A more recent study of reputation management (n = 2253) provides additional data about how OSN users manage the data they share [11]. The salient results in regard to our research include the data collected from participants in the 18-29 age group. 44% of participants in this group report they take steps to limit the amount of personal information about them online. 71% report they have changed the privacy settings on their profile to limit what they share, and 47% delete unwanted comments. These results are in contrast to the long-held belief that users do not act on their privacy concerns.

1.2 Access Control Management

Studies have found that users have a difficult time completing basic access control management tasks, including determining who has access to which resources, and making changes to an existing policy [13, 12, 10]. File sharing mechanisms tend to be so difficult to use that many users prefer to share documents as email attachments[16]. Even the designers of the Multics access control mechanisms redesigned many features after realizing the mechanisms were too complex and would lead to errors [7].

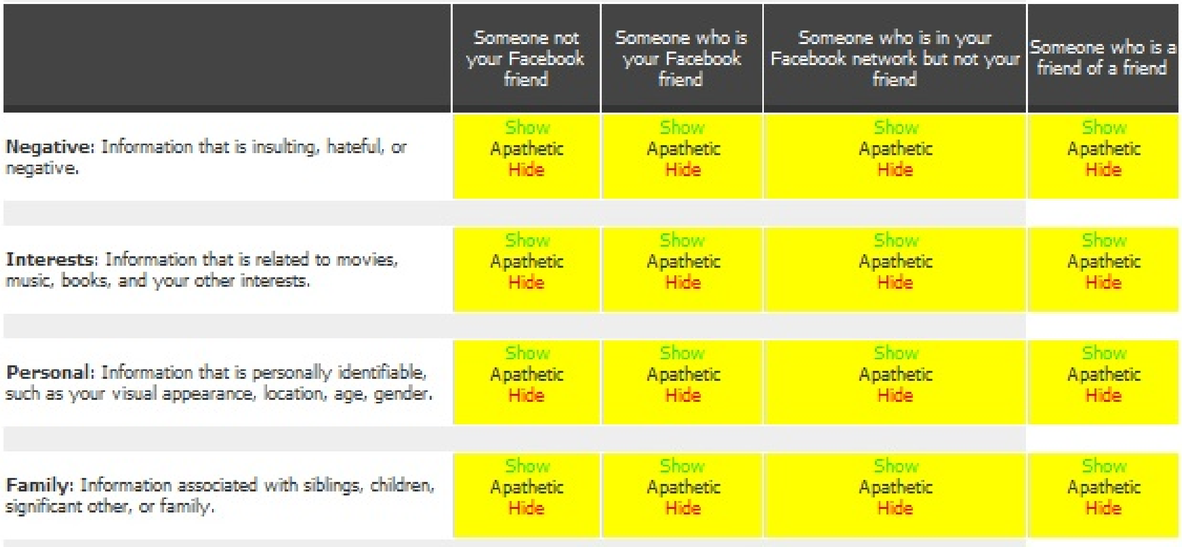
In the design of Grey, a system for physical access to rooms and computers, interviews were conducted to elicit participants' *ideal policy* for doors. Ideal policy is defined as the policy that would be enforced if the enforceable policy was not limited by implementation. One study compared participants' ideal policies to their actual policies when using physical keys, the inconsistencies motivated specific features in Grey which contributed to an improved access control mechanism where users could capture their ideal policy more closely [2].

2 Methodology

Our study investigated whether users' privacy settings match their sharing intentions; we also surveyed participant's privacy attitudes. The study was deployed using a Facebook application.[1] Participants were required to install the application, before installation they were presented with the consent form and an explanation of required permissions.

We summarize the user study stages here, provide a detailed description of each stage, and finally discuss limitations of the study design.

[1] Columbia University Protocol IRB-AAAF1543



	Someone not your Facebook friend	Someone who is your Facebook friend	Someone who is in your Facebook network but not your friend	Someone who is a friend of a friend
Negative: Information that is insulting, hateful, or negative.	Show Apathetic Hide	Show Apathetic Hide	Show Apathetic Hide	Show Apathetic Hide
Interests: Information that is related to movies, music, books, and your other interests.	Show Apathetic Hide	Show Apathetic Hide	Show Apathetic Hide	Show Apathetic Hide
Personal: Information that is personally identifiable, such as your visual appearance, location, age, gender.	Show Apathetic Hide	Show Apathetic Hide	Show Apathetic Hide	Show Apathetic Hide
Family: Information associated with siblings, children, significant other, or family.	Show Apathetic Hide	Show Apathetic Hide	Show Apathetic Hide	Show Apathetic Hide

Figure 1: The color user interface to collect participant's sharing intentions in Stage 2.

Stage 1: Survey of Privacy Attitude: Survey the participant's privacy attitudes and their experience with Facebook.

Stage 2: Collection of Intentions: Gather participant's sharing intentions for each profile group per information category using a table of information categories and profile groups.

Stage 3: Identification of Potential Violations: Examine participant's Facebook data to identify potential violations based on the intentions stated in Stage 2.

Stage 4: Confirmation of Violations: Present participant with their potential violations, allow them to confirm the actual violations, and survey their intent to act on the violation.

2.1 Stage 1: Survey of Privacy Attitude

The survey was designed to measure the user's privacy priorities, confidence in existing settings, Facebook usage, history of privacy violations, and exposure to privacy-related media coverage (see Appendix A).

2.2 Stage 2: Collection of Intentions

A user's sharing intentions, like other access control decisions, depend on the viewer and the information. For this reason, participants were presented with a table to express their sharing intentions (Figure 1). The columns displayed profile groups and the rows displayed information categories. In each cell, the participant indicated their intention for the information category and the profile group. The participants were informed that the information categories were based on content, rather than data type, and spanned all data types (*e.g.*, wall posts, photos, links, status updates). A detailed description of the profile groups and information categories follows below.

Most privacy interfaces give users two mutually exclusive options of showing or hiding information. These options, however, can not fully capture user intent. There may be information where a user does not have a strong opinion either way. For this reason, apathetic is included as an option when recording sharing intentions. Apathetic is also a useful option for our analysis because if the user views certain Facebook information as benign, feeling apathetic toward the violation may drastically affect whether the user takes corrective action. (We note that it is a truism in switching theory that increasing the number of "don't care" states improves optimization.)

2.2.1 Profile Groups

Facebook's current privacy settings display configuration options based on default profile groups (friends, friends of friends, network members, or everyone) or friend lists, which are configured manually. Our study focused on the settings for the default groups.

We redefined the default group everyone to 'stranger' to indicate that the participant had no relationship to the user (i.e. not a network member, friend of a friend, or friend). A network member is a Facebook user that shares network membership with the user. A friend of a friend is a Facebook user who is a friend of one of the user's friend. A friend is a Facebook user that the user has confirmed as a friend. In reality these groups may overlap, however, the study focused on profiles that fit in exactly one profile group.

2.2.2 Information Categories

Facebook currently permits privacy settings to be configured for basic information and for each data type. Basic information has a separate setting for each field; these fields include gender, religious views, interests, and activities. These fields tend to be fairly static; users rarely update their gender, religious views, etc. For data types (e.g., photos, notes, links, events, and status updates), which tend to be more dynamic, the user is able to select default settings per type. They can also set privacy controls per individual data object when the data is submitted.

The keyword lists were created manually, prior to recruiting, by collecting unique words that were common to the information category. The categories were selected based on themes that were most likely to elicit user intent to show or hide from certain profile groups. Sources consulted include existing Facebook data, terminology lists, and tags on related online content. The categories are not intended to be exhaustive as the study did not seek an upper bound of Facebook privacy violations. However, it is worth noting that since the participants were later prompted to verify the violation, the algorithm was liberal in its recognition of potential violations. For example, 777 objects were reported as misclassified out of a possible 3,120 combinations (4 profile groups, 12 information categories, 65 users). The following list presents the categories and gives a few sample keywords per category, many more keywords were used.

- Religious: god, priest, torah, mosque.
- Political: obama, republican, climate.
- Alcohol: drunk, beer, keg.
- Drugs: weed, smoked, tok.
- Sexual: sex, porn, hooker.
- Explicit: sh*t, f*ck.
- Academic: homework, professor, lecture.
- Work: boss, internship, interview.
- Negative: hate, sucks, ugly.
- Interests, related to interests (e.g., movies, T.V. shows): band, movie, book.
- Personal, information that is personally identifiable (e.g., location, age, gender): birthday, new york.
- Family: father, sister, mom.

2.3 Stage 3: Identification of Potential Violations

The purpose of Stage 3 was to identify all the potential violations based on the participant's sharing intentions reported in Stage 2. First, a list was compiled of all information categories where the participant indicated a show or hide intention; apathetic intentions were ignored since they cannot lead to a violation.

After compiling the list of information categories with a show or hide intention, it was necessary to identify specific objects that matched each information category. The application examined all text data associated with the participant's profile and activity. To borrow terms from Schneier's taxonomy of social networking data, the application examined disclosed, entrusted, and incidental data [14]. Disclosed data is posted by the user. Entrusted data is posted by the user on another user's profile. Incidental data is posted by a Facebook friend of the user on the user's profile. Entrusted data suggests the user has no control over the data once it's posted. On Facebook, however, the user has the option to delete the data. For this study we modify the definition of control to mean a lack of *privacy control* via privacy settings, meaning the user can *delete* the object but cannot *hide* the object. The following lists demonstrate the fields that were checked; they are not exhaustive.

- Disclosed data: status updates and their comments, photo captions, comments on photos, links, comments on links, album captions, album comments, video captions, comments on videos, notes, comments on notes, comments on wall posts, basic profile information and page memberships.
- Entrusted data: comments on friend's status updates the participant was tagged in, comments on friend's photos (and tagged photos), comments on friend's albums, comments on friend's videos (and tagged videos), comments on friend's notes (including tagged notes), comments on friend's wall, event RSVPs, public group memberships.
- Incidental data: comments on status updates, status updates the participant was tagged in, comments on friend's status updates the participant was tagged in, comments on the participant's photos and tagged photos, comments on participant's links, comments on participant's albums, comments on participant's videos, comments on participant's notes and tagged notes, comments on participant's wall.

The text of each object was searched for keywords to assign an information category. If a keyword was found in the text and the information category matched a show or hide intention provided by the user, then the ID of the object was recorded.

2.3.1 Identify Potential Violations per Profile Group

Four profiles were created to conduct the study: each configured as either a *friend*, a *friend of a friend*, a *network member*, or a *stranger*. The participant was required to accept a friend request from the *friend*. The *friend of a friend*'s only friend was *friend*. *Network member* was a member of the Columbia University network. *Stranger* did not have any friends and was not a member of any networks. Only 'network member' was a member of the Columbia University network.

In order to determine the set of potential violations it was necessary to check the privacy settings of each object of interest (an object that matched a category with a show or hide intention). This required iterating over each information category with a show or hide intention and attempting to access the object from the study profile for each profile group. Lists were created of the objects that were hidden where the category was intended to be visible, and of the objects that were visible where the category was intended to be hidden.

2.4 Stage 4: Confirmation of Violations

The purpose of Stage 4 was to allow the participant to review the potential violations and confirm the actual violations. In this stage, the participant proceeded through twelve screens, one per information category. The

Intent	Result Based on Privacy Settings	Success
Hide	At least one object matched the category and was accessible to the profile group.	No
Hide	All objects that matched the category were hidden from the profile group.	Yes
Show	At least one object matched category and was hidden to the profile group.	No
Show	All objects that matched the category were accessible to the profile group.	Yes

Table 1: Possible Results

	18	19	20	21	22	23	24	25	
Female	2	3	8	8	5	10	3	1	40
Male	3	2	7	4	3	2	1	3	25
	5	5	15	12	8	12	4	4	65

Table 2: Participant Demographics

screen was divided into four sections, one per profile group. Each section displayed the participant's sharing intention for a profile group and presented information to represent one of four possible states presented in Table 1.

In Table 1, the rows where success is 'No' indicate a *potential* privacy violation, these cases may represent instances where the participant's privacy settings do not match their reported sharing intentions. We define a *hide violation* to be the case where the intent was to hide for the profile group, but one or more objects in the category was accessible. We define a *show violation* to be the case where the intent was to show for the profile group, but one or more objects in the category was not accessible.

For potential violations, the application retrieved the object in question and displayed it to the participant. The justification (i.e. the found keyword) for the potential violation was shown in boldface to provide the participant with context. Within each section the potential violations were grouped based on the source (whether the data was posted by the participant or a friend) and on the data type (photo comment, group, event, status update, etc). Participants were asked to confirm which potential violations were real; it is not possible to distinguish an actual violation from a potential violation without input from the user.

Also of interest was whether the violation motivated the participant to modify past settings or future behavior. Since the participant was often presented with several pieces of information, they were instructed to answer the questions *only* on correctly classified information, and to ignore misclassified information. This procedure, although limited, was optimal for the study's goal. The nature of misclassified information was not investigated further, this study focused strictly on the extent and severity (albeit lower bound) of sharing violations.

2.5 Participants

Participation was limited to Columbia University students. This limitation was required in order for the application to measure the participant's privacy settings for network members (one of the four default groups). 65 participants were recruited on the Columbia University campus. Age and gender demographics of the participants are provided in Table 2. Recruitment methods included flyers, broadcasts to Facebook groups, broadcasts on mailing lists, and a paid advertisement on a campus blog. The recruiting material

	mean	med	s. dev.	max	min
None posted	3.06	3	1.04	4	1
Never untag	2.94	3	1.01	4	1
Physical security	3.00	3	.77	4	1
Unattractive likeness	1.98	2	.89	4	1
Undesired behavior	2.22	2	.84	4	1

Table 3: Responses for Q7 on Photo Untagging Behavior (1 = most common).

targeted the student body at large. Thus, the final sample was a convenience sample of students who responded to the advertisements. Participants were compensated $10.

2.6 Limitations of Study Design

The Facebook API does not allow applications to directly query the privacy settings of a user. To work around this issue we chose to focus on checking the privacy settings for the four default profile groups since it is simple to create a new profile and require participants to accept a friend request, create a second profile as a friend of the first, and create a third profile to represent a stranger. In theory, the only profile that would be difficult to create is a network member. However, in our case this was trivial given our association with Columbia University. The necessity of having a profile in the same network restricted our recruiting to Columbia University students.

Typically, sampling only students can be a weakness; for this study, it may be an advantage, since students are generally quite tech-savvy, and (based on their age) are almost certainly experienced Facebook users. This suggests that if any subset of users would be adept at managing privacy settings it might be the one we surveyed. They are also likely to have more potentially sensitive information posted online compared to older demographics, and will be on the job market soon meaning correct use of privacy settings is quite important.

In regard to the participant sample, the prerequisite of installing a Facebook application and granting full offline access to the study application may have biased the sample. During a pilot study we received several comments from potential participants who opted-out because of this requirement. We did not, however, receive similar feedback when recruiting on campus.

The ideal sampling of Facebook users would require knowing the demographics on usage habits of Facebook's user base which, at this scale, is information that only Facebook has access to. Given the unequivocal nature of many of the results reported in this paper, the sample size and sampling size are adequate to provide initial insight to the trouble users experience with the existing privacy settings.

3 Results

In this section we present results from each stage of the user study, and discuss correlations between stated privacy priorities, sharing intentions, and responses to the actual violations.

3.1 Stage 1: Privacy Attitudes

In question 1, participants were asked to choose a response to represent the most important reason for online privacy. 49% of the participants selected reputation security (protecting social reputation), 38% selected economic security (preventing identity theft and protecting browsing habits from advertisers and third parties), and the remaining 12% selected physical security (to protect me physically, by hiding my face, location, and/or contact information from strangers).

Questions 2-4 asked the participant to express their level of concern with economic, reputation, and physical security options ranged from 'why would I be concerned?' to 'I'm very concerned'. The possible responses were converted to a numerical score from 1 to 5 (very concerned). Selecting each participant's

	Friend Count	Economic	Reputation	Physical	7a	7b	7c	7d	7e
Friend Count	1								
Economic	-.23	1							
Reputation	.01	.27	1						
Physical	-.11	.21	.25	1					
7a	-.07	-.06	-.20	-.03	1				
7b	-.09	-.22	-.00	-.05	.11	1			
7c	.22	-.10	-.19	-.32	.45	.18	1		
7d	.26	-.10	-.10	-.22	.05	-.38	.20	1	
7e	-.13	.14	-.20	-.02	.20	-.13	.19	.28	1

Table 4: Pearson's Correlations of Untagging Behavior (Q7) and Security Concerns (Q2-4).

greatest response to each question (mean = 4.43) suggests that most participants are concerned with some form of privacy, and the average of their lowest response across the three questions (mean = 3.02) indicates that all aspects of privacy are of at least some concern. The correlations between all three types of security (Table 4) also shows that participant concern tends to increase collectively rather than focusing only on one security concern.

Table 4 includes correlations on friend count (mean = 601, median = 565, max = 1894, min = 160). The average friend count is much higher than the average reported by Facebook (130). The wide variance in friend count may reflect a difference in not only friending habits but also potentially the age of the Facebook account. The consistent high concern for reputation security across participants regardless of the number of friends further supports the universality of reputation management. In contrast, the negative correlations in both economic and physical security may suggest that this security is of lower concern because the participant has already implemented prudent friending habits in the belief that this behavior reduces their risk.

3.1.1 Facebook Expectations and Interactions

An overwhelming majority of participants (62 of the 65) believed their settings matched their attitude toward privacy (Q5), which is contrary to the results gathered on sharing violations. The three participants who believed their settings were incorrect each reported a different reason: not enough time, did not know how, and Facebook did not have the privacy controls they wanted (Q6).

Question 7 asked participants to rank their reason for untagging photos from 1-4 (1 = most common). Results are presented in Table 3 (lower numbers represents greater significance) and correlations to the previous questions are in Table 4; the full text of question 7 is available in Appendix A. In the table of correlations, due to the phrasing of the questions, *a negative correlation between a security concern and an response for question 7 means that the reason increases in significance as the security concern increases.* For example, the negative correlation between physical security and untagging photos for physical security (7c) shows that untagging photos for physical security increased as concern for physical security increased, further supporting that participants concerned with physical security take corrective action to minimize risk. Untagging unattractive photos, on the other hand, was universally common with a positive correlation with a participant's friend count.

In question 8, participants were asked the reason they chose to hide basic factual data on their profile. The most common reason for hiding factual data (selected by 21 participants) was due to physical security and the fear of potential harm. The repeated self-reported behavior for physical security potentially contradicts the first survey question where participants prioritized physical security lower than other types of security. However, it may simply suggest either that users are more vigilant in protecting their basic factual data and have less concern for physical harm, or basic factual data comprises a very limited subset of Facebook's total user privacy settings.

Responses to question 9 are represented in Figure 2. Each chart represents the distribution of profiles for

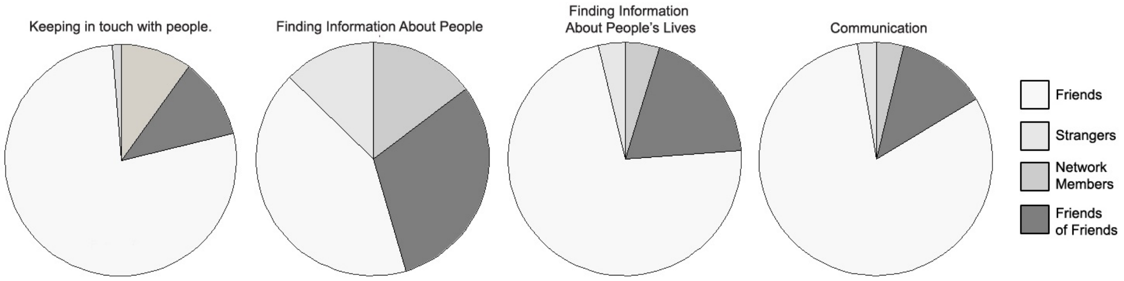


Figure 2: Typical Facebook interactions with each profile group.

	Show	Apathetic	Hide
Friend	389	201	190
Network Member	152	209	419
Friend of Friend	155	228	397
Stranger	89	166	525

Table 5: Sharing intentions for each profile group

that type of interaction. Friends are overwhelmingly the largest target for most interactions and strangers consistently represent the smallest target. Not surprisingly, strangers represent a larger portion of interactions for finding information.

In question 10, nearly all participants reported never experiencing a Facebook related information leak that had a negative impact. Less than a third of the participants reported that their privacy behavior has not been affected at all by the media coverage of Facebook privacy. About two thirds of participants reported that media coverage led them to double-check their privacy settings but did not modify anything. Of the participants who double checked their privacy settings, most also stated that they would become more selective regarding future information shared on Facebook.

3.2 Privacy Intentions

User intent and profile groups were converted into numerical scores for the purpose of analysis. The intent to hide, apathy, and to show were assigned the values of -1, 0, and 1, respectively. Profile groups were ranked from 1 to 4 based on their estimated size with 4 denoting the largest group. The friend profile group had an average size of 601, based on the friend counts of our participants. The network group size was 75,979 members, as reported by Facebook. The stranger group size was estimated at 500 million, the number of active Facebook users. Friend of friend size was estimated as 354,025, based on the average friend count.

The distributions of sharing intentions versus profile groups are displayed in Table 5 and Figure 3. There is a strong negative correlation between the size of the profile group and the intention to show. As the profile group grew in size, significantly less information was intended to be shown. Certain information categories were also consistently selected as hidden, such as drug or negative. This consistency, however, is not observed in intentions to show information suggesting that there is a consensus only on what is inappropriate on a Facebook profile.

Correlations across information categories are available in Table 7, 8, 9, and 10 (in the appendix). Certain categories exhibit high correlations, suggesting they are either similar or their privacy intentions are similar. For example explicit, sexual, political, alcohol, and religious had strong positive correlations across all profile groups. Similarly, personal, family, academic, work, and interests had high correlations across all profile groups.

The concern for physical and economic security correlated positively with selection of information categories to hide and correlated negatively with selection of information categories to show. For example,

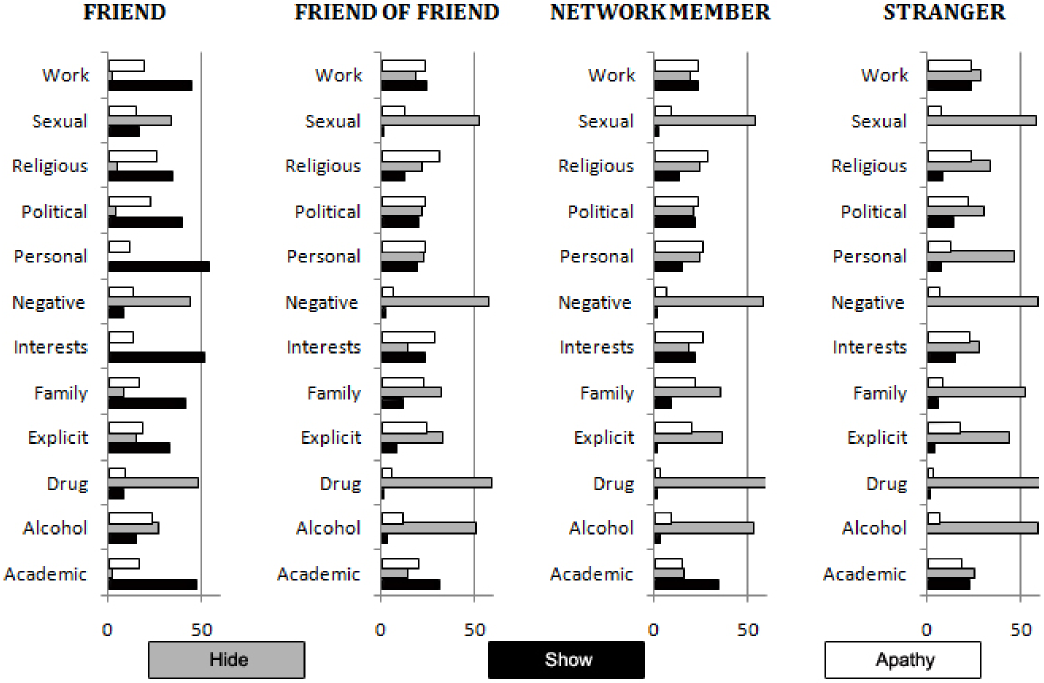


Figure 3: Sharing Intentions for each profile group

	Show Violation	Hide Violation
Nothing Modified	36	25
Only Future Settings Modified	9	14
Past and Future Modified	10	22
Total	55	61

Table 6: Greatest Response to a Violation

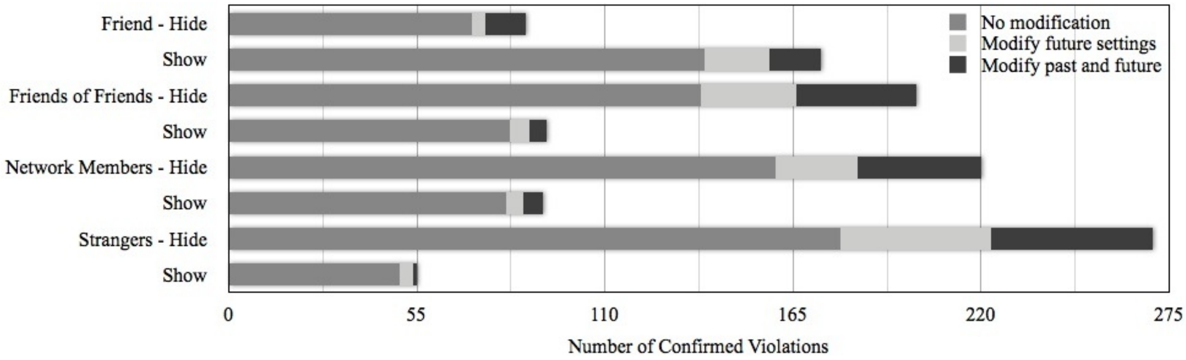


Figure 4: Number of violations and reported action by profile group.

regarding information categories to show friends, users with little economic concern selected an average of 6.36 information categories, users with economic concern selected an average of 5.97 information categories, and users with high economic concern selected an average of 5.17 information categories. This trend of increased concern correlating to decreased intention to show and increased intention to hide was observed by both physical and economic security and in all four profile groups. The observance of these trends suggests that concern for aspects security are indeed present in the user during setting configurations (albeit to arguable degree).

Concern for reputation security, however, exhibited no visible trend and the extrema of per user average selection for information categories often occurred with ‘general’ concern (rather than the expected ‘little’ or ‘high’ concern). Although this lack of correlation may be counterintuitive, the results allude to the dichotomy facing the user regarding their behavior and reputation security: is it more secure to follow a traditional model of security and attempt to hide information or should a user display information in an attempt to control and shape their reputation?

3.3 Violations

For the purpose of this study, a hide violation is visible information that was intended to be hidden to that profile group while a show violation is invisible information that was intended to be shown. Participants were first shown their potential violations for each category, then asked to confirm the actual violations. The participant was also asked how the violation might influence their behavior, past and future; their reaction to the violation may indicate the severity of the violation.

Our results show that there is serious potential for privacy problems. 93.8% of participants revealed some information that they did not want disclosed. Given our sample, it is virtually certain that most other Facebook users have similar problems. On the other hand, we also note that 84.6% of participants are hiding information that they wish to share. In other words, the user interface design is working against the very purpose of online social networking. Between these two figures, every single participant had at least one sharing violation. Either result alone would justify changes; taken together, the case is quite compelling.

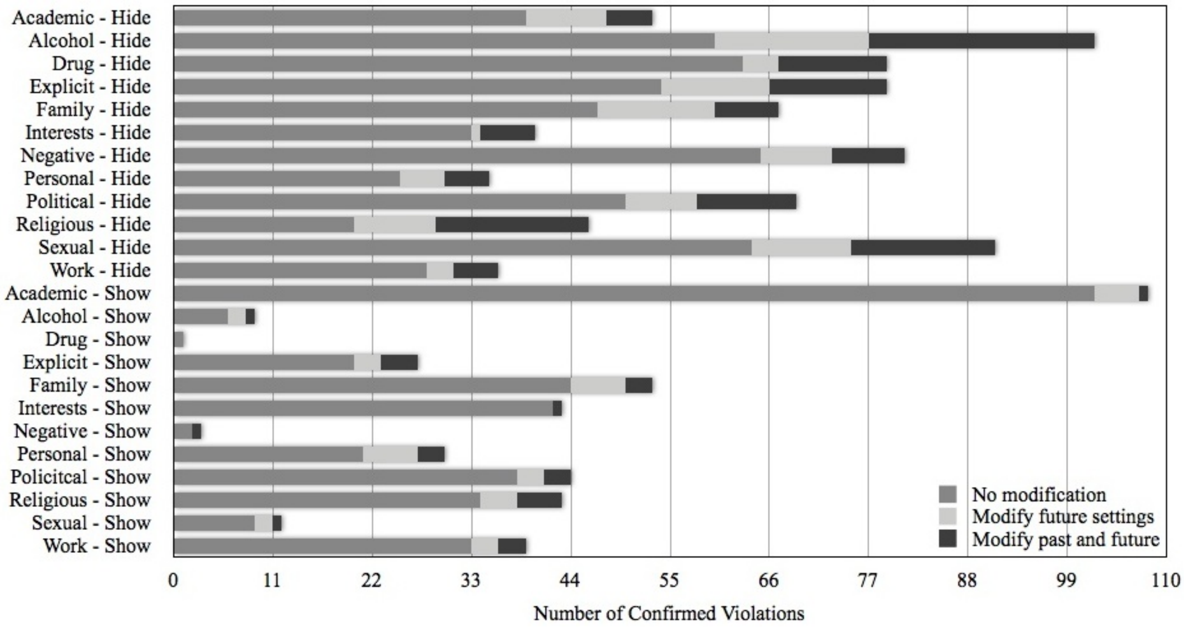


Figure 5: Number of violations and reported action by category.

For each confirmed violation, the participant was asked to state the influence the violation might have on their behavior using one of the following options.

1. Accurate, but I will not modify anything.

2. Accurate, I will be more attentive to future settings.

3. Accurate, I will be more attentive to future settings and will modify past settings.

The response to this question was used to gauge the severity of the violation as gauged by the participant, as this will vary by individual. In table 6 each participant's greatest response to a confirmed violation was tallied (using the choices listed previously). The numbers in the table are indicative of the overall adequacy of the participant's privacy settings based on their sharing intentions, and our data categories. Many participants' worst confirmed violation was not serious enough to motivate a change in behavior. Though approximately one third of the participants would like to change their past and future settings to correct at least one violation. For both types of violations the most common response was 'no modification' which could mean that the error is not worth correcting, is not that serious, or that the participant does not know how to address it.

Figure 4 presents the number of confirmed hide and show violations per profile group. Each bar is further divided based on the reaction to the violation. The data indicate that a show violation is more likely to motivate a behavior change when the information is hidden from a friend (34 of 173 friend, show violations) The other three profile groups also had far fewer show violations. The stranger group had the most hide violations and the most violations that would influence past and future settings. Overall, hide violations were most likely to influence future behavior, especially for the non-friend groups (friends of friends = 35, network member = 36, and stranger = 47).

Figure 5 presents the number of confirmed violations per information category. The information categories most likely to motivate a change in future and past settings are alcohol, drug, explicit, sexual, and religious. The disproportionate number of academic show violations may be an artifact of the participant sample.

4 Recommendations for Facebook

Every single participant in the study confirmed that at least one of the potential violations contradicted their sharing intentions. Although not all contradictions elicited corrective action, the existence of these violations presents a clear message: not only are Facebook's existing privacy paradigms flawed but immediate improvements must be made to minimize risk to users.

4.1 Contextual Privacy Settings

One of the largest culprits for privacy flaws is Facebook's reliance on data types for defining default privacy settings. These datatypes are misrepresentative of the real world that Facebook attempts to model. Outside of a social network, an individual does not determine visibility of personal data by its format but instead by the context of its information. A key privacy improvement by Facebook, therefore, would be to automatically categorize information with a predicted context, each which would have its own default privacy configuration reflecting the user's intent with the data. The study has already shown that users are strongly opinionated in either showing or hiding specific categories of information. With these intentions serving as default privacy settings, the existence of these information categories may have not only caught most of the privacy violations found by the study but also provides a useful representation for users regarding the nature of their Facebook data.

Although Facebook technically already allows for privacy configuration on most data at an individual level (e.g. each photo album, video, note, status update, etc), users have such a large amount of data that even if they regularly utilized this feature their usage would be subject to several oversights. Given the large amount of data to manage, the usability of an information categorization feature may, therefore, also benefit significantly from automation. The study had success in automatically identifying information categories using a primitive text search algorithm. More advanced approaches such as machine learning, natural language processing, or image analysis techniques may address the liberal nature of the algorithm and drastically improve usability for configuring privacy settings on existing and future Facebook data. The ability to tag new data items with an information category can improve methods that attempt to learn user privacy policies, which can help relieve some of the burden of utilizing fine-grained access control [8].

Information categorization also provides extensibility for future features. Currently, new features, such as Facebook Places, are implemented by introducing entirely new privacy settings and have historically required the user to opt-out of the new feature to maintain their previous level of privacy. Users (and the media) rarely approve of such impositions on data visibility. Our suggested information tags feature provides an implementationally feasible way to derive settings for unanticipated features. For example, in the case of Facebook Places, if a user wishes to hide alcohol related information from everyone, it is reasonable to conclude that all location check-ins at a bar should be hidden from everyone. This method of inference provides a reasonable compromise between user privacy intentions and Facebook's bias toward data visibility. Prior work has explored the possibility of using content-based access control for blog posts, further investigation is necessary to determine if a similar approach can be used for OSN posts [5].

4.2 Private Information

Our participants overwhelmingly wished to show personal information to friends but hide it from strangers. This volatile user intent for personal information suggests not only that policies regarding personal information merit further examination but also that modifications to existing settings and future settings on behalf of the user should err on the side of restricting information visiibility. Visibility limitations would not affect the user experience since the study also showed that users have little purpose for strangers in their OSN interactions.

The study also observed less volatile information that was consistently chosen to be private for certain profile groups. Privacy violations may be significantly reduced by examining visible user data for commonly hidden information categories to check for potential oversight.

5 Future Work

Our study investigated users' sharing intentions and actual privacy settings in search of violations. The fact that every participant confirmed at least one sharing violation indicates that additional research on the usability of privacy settings is necessary. Determining the root cause of violations is one possible follow-up study; this is better suited to an in-person interview (as opposed to the remote study reported here), this would allow study coordinators to adjust the questions to identify the source of the violation. Participants who have violations may not understand the privacy settings well enough to identify the reason behind a violation in any format but an interview.

Our participants were shown potential violations and asked to indicate which were actual violations and how their future behavior may be impacted by the type of violation. A number of them claimed they would modify existing settings, or change their future behavior. Another follow-up study would involve querying those users who claimed they would modify their behavior and evaluating whether they actually did. When it appeared that they did not alter their behavior, it is important to determine why. Were they unable to modify their privacy settings to address the violation, or were they insufficiently motivated to change their settings? The results of such a study could influence the design of privacy settings.

The results of our study suggest ternary privacy settings may be useful. Participants were given a sharing choice of show, hide, or apathetic for each information type and profile group pair. The participants selected apathetic more often than anticipated. The accuracy of privacy settings may be improved if users were given an apathetic or 'don't care' option. Future work should further investigate the implications of an apathetic option.

6 Conclusion

We conducted an empirical evaluation of actual preferences and behavior of Facebook users. Every one of our 65 participants had at least one sharing violation based on their stated sharing intentions. In other words, every participant was sharing something they wished to hide, or was hiding something they wished to share. Both cases represent a shortcoming of the privacy settings. Our results provide a lower bound of the inconsistencies between users' sharing intentions and their privacy settings. We were limited by the Facebook API and potentially by our sample of Facebook users. However, the results still indicate that the current approach is fundamentally flawed.

We recommend improvements to the current mechanism based on our findings, and suggest directions for future work. We also note that as outside researchers we have a limited ability to perform these studies. To obtain more accurate results we recommend Facebook regularly conduct similar studies to evaluate their privacy mechanisms, especially as new features are introduced. An earlier version of this paper included an additional recommendation based on our findings that friend-contributed data was a leading culprit of violations. We deleted the recommendation when we noticed Facebook has already augmented their privacy settings to include this feature. We applaud that change, but feel that much more could and should be done.

References

[1] Acquisti, A., and Gross, R. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Privacy Enhancing Technologies*. Springer Berlin, 2006, pp. 36–58.

[2] Bauer, L., Cranor, L. F., Reeder, R. W., Reiter, M. K., and Vaniea, K. A user study of policy creation in a flexible access-control system. In *CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems* (NY, NY, USA, 2008), ACM, pp. 543–552.

[3] Besmer, A., and Richter Lipford, H. Moving beyond untagging: photo privacy in a tagged world. In *CHI '10: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2010), ACM, pp. 1563–1572.

[4] Gross, R., and Acquisti, A. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society* (New York, NY, USA, 2005), WPES '05, ACM, pp. 71–80.

[5] Hart, M., Johnson, R., and Stent, A. More content-less control: Access control in the web 2.0. In *WOSP '08: Proceedings of the first workshop on Online social networks* (2008), pp. 43–48.

[6] Joinson, A. N. Looking at, looking up or keeping up with people?: motives and use of Facebook. In *CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2008), ACM, pp. 1027–1036.

[7] Karger, P. Personal communication, Apr. 2009.

[8] Kelley, P. G., Hankes Drielsma, P., Sadeh, N., and Cranor, L. F. User-controllable learning of security and privacy policies. In *Proceedings of the 1st ACM workshop on Workshop on AISec* (New York, NY, USA, 2008), ACM, pp. 11–18.

[9] Kumaraguru, P., and Cranor, L. F. Privacy indexes: A survey of Westin's studies. Tech. Rep. CMU-ISRI-5-138, Carnegie Mellon University, 2005.

[10] Lipford, H. R., Besmer, A., and Watson, J. Understanding privacy settings in Facebook with an audience view. In *UPSEC'08: Proceedings of the 1st Conference on Usability, Psychology, and Security* (Berkeley, CA, USA, 2008), pp. 1–8.

[11] Madden, M., and Smith, A. Reputation management and social media. `http://pewinternet.org/Reports/2010/Reputation-Management.aspx`, May 2010.

[12] Reeder, R. W., Kelley, P. G., McDonald, A. M., and Cranor, L. F. A user study of the expandable grid applied to P3P privacy policy visualization. In *WPES '08: Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society* (New York, NY, USA, 2008), ACM, pp. 45–54.

[13] Reeder, R. W., and Maxion, R. A. User interface dependability through goal-error prevention. *International Conference on Dependable Systems and Networks* (2005), 60–69.

[14] Schneier, B. A taxonomy of social networking data. *IEEE Security and Privacy 8* (2010), 88.

[15] Thomas, K., Grier, C., and Nicol, D. M. unFriendly: Multi-party privacy risks in social networks. In *Proceedings of the 10th international conference on Privacy enhancing technologies* (2010), Springer-Verlag, pp. 236–252.

[16] Whalen, T., Smetters, D., and Churchill, E. F. User experiences with sharing and access control. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems* (New York, NY, USA, 2006), ACM, pp. 1517–1522.

A Privacy Attitude Survey

1. Choose one of the following to represent what you believe is the most important reason for online privacy.

   (a) Economic Security: To prevent identity theft and protecting browsing habits from advertisers and third parties.

   (b) Reputation Security: To hide information to protect my social reputation.

   (c) Physical Security: To protect me physically, by hiding my face, location, and/or contact information from strangers.

2. Are you concerned with online privacy as related to economic security?
    (a) Why would I be concerned?
    (b) I'm not concerned
    (c) I'm a little concerned
    (d) I am concerned
    (e) I'm very concerned
3. Are you concerned with online privacy as related to reputation security?
    (a) Why would I be concerned?
    (b) I'm not concerned
    (c) I'm a little concerned
    (d) I am concerned
    (e) I'm very concerned
4. Are you concerned with online privacy as related to physical security?
    (a) Why would I be concerned?
    (b) I'm not concerned
    (c) I'm a little concerned
    (d) I am concerned
    (e) I'm very concerned
5. Do you feel your Facebook settings reflect your attitude related to privacy?
    (a) Yes.
    (b) No.
    (c) I am not concerned with privacy.
6. (Shown if response to Q5 is 'No' or 'I am not concerned with privacy').
    (a) I know how to change the settings but I don't have the time to do it.
    (b) I don't know how to change the settings.
    (c) Facebook doesn't have the privacy controls I want.
7. For which reasons have you untagged (not HIDDEN) photos? Rank with the lowest number marking the most common reason.
    (a) I have chosen not to post any photos.
    (b) I have never untagged a photo.
    (c) The photo displayed my face or location, which I have chosen to keep secret to protect my physical security.
    (d) I didn't like the photo of me (it was unattractive or not flattering).
    (e) The photo displayed behavior I did not want to be associated with (something that could be embarrassing if others saw it).
8. For which reasons have you hid posted factual data (ex. Birthday, hometown, gender, etc)?

(a) The information could potentially be used for identity theft.

(b) I do not feel safe with that information out there since I believe I could be potentially stalked, found, and/or harmed.

(c) I don't want other people to know how old I am, or where I am from, for professional or social reasons.

(d) I have not hid any information.

(e) I have chosen not to enter factual data on my profile.

9. Why do you use Facebook? Check all that apply for the respective groups: friends, friends of friends, network members, and strangers.

    (a) Keeping in touch with people.

    (b) Finding information about people (i.e. profile watching).

    (c) Finding information on people's daily lives (i.e. newsfeed, status updates)

    (d) Communication (i.e. messages, wall posts, etc).

10. Have you ever had an accidental leak of information on Facebook that had a negative impact? If so, what happened?

    (a) I never had an accidental leak of information.

    (b) I was a victim of identity theft or an online account was hacked into.

    (c) I was physically harmed, stalked, or contacted by someone due to the release of information.

    (d) Information sensitive to my social reputation was viewed by a friend or coworker who I did not want see the information.

11. What type of information was accidentally leaked?

    (a) Fact-type information (gender, birthday) available on my profile.

    (b) Activity (photo, status update) information posted by me.

    (c) Activity (photo, status update) information posted by others.

    (d) I did not have a privacy misstep.

12. What do you think was the cause of the information leak?

    (a) I'm not sure. I didn't think the person would be able to view it based on my privacy settings.

    (b) I had not changed any of my privacy settings.

    (c) I had not changed my privacy settings for that type of information.

    (d) I didn't remember that information was on my profile.

    (e) I didn't expect my friend to post that information on Facebook.

13. If you suffered from a privacy misstep, did you alter your behavior? How?

    (a) No information leak.

    (b) I did not alter my behavior.

    (c) I became more selective about the information I put on Facebook.

    (d) I disabled some features of Facebook, like my wall.

    (e) I made changes to my privacy settings so that it wouldn't happen again.

    (f) I deleted the piece of information.

(g) I deleted that friend or put them on a limited profile view.

(h) I paid closer attention to privacy on Facebook on a per-activity basis, such as modifying the privacy level of each individual status update.

14. Have you heard anything regarding Facebook privacy in the news lately?

(a) I haven't heard anything.

(b) I haven't heard anything but, then again, I don't really read the news.

(c) I read a headline or heard something briefly I didn't really care to investigate further.

(d) I heard something about Facebook and it seemed negative but I don't know any further details.

(e) I heard something about Facebook and it seemed positive but I don't know any further details.

(f) Facebook released my private information to advertisers.

(g) Facebook released my private information to the general public.

(h) Facebook has released an improved privacy interface.

(i) Facebook has released a new privacy interface.

15. Where did you hear it from? Check all that apply.

(a) Somebody told me in person or over the phone.

(b) General News Source

(c) On Facebook (such as from a friend's status)

(d) Privacy-related source.

16. Has the media affected your behavior on Facebook?

(a) It has not affected my behavior at all.

(b) I became more selective about the information I post on Facebook.

(c) I deleted some Facebook friends.

(d) I modified my privacy settings.

(e) I double-checked my privacy settings but did not change them.

	S.	Pe.	F.	E.	D.	Ac.	W.	Po.	Al.	N.	I.	R.
Sexual	1.00											
Personal	0.29	1.00										
Family	0.29	0.44	1.00									
Explicit	0.47	0.35	0.36	1.00								
Drug	0.32	0.25	0.29	0.26	1.00							
Academic	0.12	0.36	0.05	0.19	0.07	1.00						
Work	0.33	0.39	0.03	0.39	0.32	0.38	1.00					
Political	0.45	0.20	0.33	0.28	0.24	0.13	0.44	1.00				
Alcohol	0.44	0.26	0.39	0.32	0.56	0.05	0.17	0.24	1.00			
Negative	0.47	0.17	0.10	0.44	0.25	0.25	0.13	0.23	0.46	1.00		
Interests	0.36	0.49	0.22	0.33	0.28	0.15	0.36	0.34	0.23	0.18	1.00	
Religious	0.41	0.25	0.25	0.25	0.24	0.14	0.42	0.73	0.29	0.21	0.25	1.00

Table 7: Information Category Correlations for Friends

	S.	Pe.	F.	E.	D.	Ac.	W.	Po.	Al.	N.	I.	R.
Sexual	1.00											
Personal	0.51	1.00										
Family	0.44	0.61	1.00									
Explicit	0.56	0.57	0.44	1.00								
Drug	0.24	0.02	0.13	0.23	1.00							
Academic	0.23	0.46	0.35	0.41	0.01	1.00						
Work	0.29	0.42	0.31	0.34	-0.03	0.55	1.00					
Political	0.28	0.52	0.57	0.56	0.23	0.45	0.39	1.00				
Alcohol	0.54	0.36	0.48	0.48	0.26	0.20	0.16	0.38	1.00			
Negative	0.70	0.42	0.39	0.50	0.29	0.24	0.22	0.28	0.49	1.00		
Interests	0.28	0.45	0.53	0.46	0.18	0.41	0.39	0.60	0.26	0.36	1.00	
Religious	0.40	0.47	0.55	0.47	0.31	0.46	0.38	0.72	0.31	0.37	0.54	1.00

Table 8: Information Category Correlations for Friends of Friends

	S.	Pe.	F.	E.	D.	Ac.	W.	Po.	Al.	N.	I.	R.
Sexual	1.00											
Personal	0.42	1.00										
Family	0.42	0.52	1.00									
Explicit	0.65	0.54	0.33	1.00								
Drug	0.10	-0.08	0.00	0.07	1.00							
Academic	0.25	0.40	0.29	0.42	-0.08	1.00						
Work	0.25	0.49	0.36	0.36	-0.14	0.62	1.00					
Political	0.33	0.45	0.37	0.45	0.12	0.46	0.38	1.00				
Alcohol	0.57	0.39	0.49	0.54	0.08	0.21	0.26	0.26	1.00			
Negative	0.65	0.39	0.24	0.42	0.05	0.19	0.18	0.21	0.49	1.00		
Interests	0.40	0.45	0.39	0.45	-0.01	0.43	0.47	0.50	0.32	0.35	1.00	
Religious	0.36	0.37	0.34	0.36	0.19	0.45	0.44	0.65	0.26	0.30	0.47	1.00

Table 9: Information Category Correlations for Network Members

	S.	Pe.	F.	E.	D.	Ac.	W.	Po.	Al.	N.	I.	R.
Sexual	1.00											
Personal	0.31	1.00										
Family	0.17	0.53	1.00									
Explicit	0.52	0.23	0.13	1.00								
Drug	0.23	-0.07	-0.03	0.17	1.00							
Academic	0.25	0.46	0.33	0.46	0.01	1.00						
Work	0.18	0.49	0.28	0.34	-0.05	0.69	1.00					
Political	0.30	0.33	0.38	0.50	0.20	0.51	0.51	1.00				
Alcohol	0.58	0.05	0.12	0.50	0.59	0.21	0.03	0.30	1.00			
Negative	0.58	0.20	0.12	0.50	0.26	0.14	0.10	0.24	0.45	1.00		
Interests	0.28	0.41	0.41	0.46	0.06	0.45	0.48	0.59	0.22	0.28	1.00	
Religious	0.41	0.31	0.42	0.40	0.28	0.54	0.51	0.72	0.41	0.34	0.47	1.00

Table 10: Information Category Correlations for Strangers