Moving Beyond Set-It-And-Forget-It Privacy Settings on Social Media

Mainack Mondal
IIT Kharagpur / University of Chicago
mainack@cse.iitkgp.ac.in

Günce Su Yılmaz
University of Chicago
suyilmaz@uchicago.edu

Noah Hirsch
University of Chicago
nashirsch@uchicago.edu

Mohammad Taha Khan
University of Illinois at Chicago
taha@cs.uic.edu

Michael Tang
University of Chicago
mtang72@uchicago.edu

Christopher Tran
University of Illinois at Chicago
ctran29@uic.edu

Chris Kanich
University of Illinois at Chicago
ckanich@uic.edu

Blase Ur
University of Chicago
blase@uchicago.edu

Elena Zheleva
University of Illinois at Chicago
ezheleva@uic.edu

ABSTRACT

When users post on social media, they protect their privacy by choosing an access control setting that is rarely revisited. Changes in users' lives and relationships, as well as social media platforms themselves, can cause mismatches between a post's active privacy setting and the desired setting. The importance of managing this setting combined with the high volume of potential friend-post pairs needing evaluation necessitate a semi-automated approach. We attack this problem through a combination of a user study and the development of automated inference of potentially mismatched privacy settings. A total of 78 Facebook users reevaluated the privacy settings for five of their Facebook posts, also indicating whether a selection of friends should be able to access each post. They also explained their decision. With this user data, we designed a classifier to identify posts with currently incorrect sharing settings. This classifier shows a 317% improvement over a baseline classifier based on friend interaction. We also find that many of the most useful features can be collected without user intervention, and we identify directions for improving the classifier's accuracy.

CCS CONCEPTS

• **Security and privacy → Usability in security and privacy**.

KEYWORDS

privacy settings, access control, retrospective privacy, predictor

ACM Reference Format:
Mainack Mondal, Günce Su Yılmaz, Noah Hirsch, Mohammad Taha Khan, Michael Tang, Christopher Tran, Chris Kanich, Blase Ur, and Elena Zheleva. 2019. Moving Beyond Set-It-And-Forget-It Privacy Settings on Social Media. In *2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19), November 11–15, 2019, London, United Kingdom.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3319535.3354202

CCS '19, November 11–15, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6747-9/19/11.
https://doi.org/10.1145/3319535.3354202

1 INTRODUCTION

For users to select a proper access control setting when sharing data, they must consider the intended audience, their personal preference, and the broader context. In many cases, this decision is "set it, and forget it." That is, the access control decision made when initially sharing the data persists until it is changed, even if a user would no longer choose that same setting. For instance, a Facebook post made in 2006 when a user was a college student with 100 Facebook friends may have very different implications in 2019 when the user is a parent in the workforce with 2,000 Facebook friends. Whether a privacy setting chosen in 2006 still applies in 2019 could depend on an innumerable collection of potential features, which points to a significant cognitive burden for users. This burden is compounded by the sheer volume of posts accumulating over time, all in need of privacy setting reevaluation. As a result, manual retrospective privacy management is nigh impossible for users.

While a manual approach is completely untenable, many of the potentially predictive features that could help automate this process are personal enough that the only way to understand them is to ask the user. As a result, the explosion of potential features calls for a system design strategy which incorporates deep user interaction into an iterative, breadth-first approach. The questions asked during these interactions should be driven by hypotheses about potentially effective inference and should be consistent with the combined goal of minimizing incorrect privacy settings, not interrupting users, and minimizing data collection.

To understand risks posed by shifting privacy preferences and to identify features that could be used to identify potentially incorrect settings, we conducted a study of 78 Facebook users. With each participant's informed consent, we automatically analyzed their full timeline and activity log. We then asked participants to reevaluate five posts' privacy settings and indicate whether six chosen Facebook friends should be able to access each post. While prior studies have used the Facebook API in concert with user surveys to evaluate Facebook privacy settings [4, 6, 23, 33], to our knowledge we are the first to evaluate these privacy settings contextualized in an account's full history, including changes in friends over time.

Our participants were active Facebook users, and 71% of them had accounts that were at least a decade old, providing a rare look

into the longitudinal evolution of Facebook privacy settings. In contrast with prior longitudinal work on Facebook privacy [47, 73], we found that participants' most common privacy setting was "Friends only." We also note that the median participant had four times as many friends in 2018 as they did in 2009. The meaning of the "Friends only" setting, and thus the visibility of such posts, has changed substantially over time. Participants expected their Facebook friends to sometimes look at their old posts, emphasizing the importance of updating privacy settings even for old content.

While 45% of participants reported having used Facebook's "privacy checkup," current retrospective privacy management mechanisms are insufficient. A number of the privacy settings active on participants' Facebook accounts did not reflect their current intentions. Overall, 65.3% of participants reported wanting to change the privacy setting of at least one of the five posts we presented to them. This represented 25.5% of posts participants saw, with rough parity between increasing and decreasing visibility.

Using insights from the user study regarding how users conceptualize and decide to modify their privacy settings, we built models to predict which posts from the history of a user's Facebook account are most likely to have active privacy settings that no longer match the user's intent, as well as which posts perhaps should not be shared with specific Facebook friends. Due to the sensitive and subjective nature of managing privacy settings, we optimized our prediction algorithm design for deployment as part of a human-in-the-loop model that augments, rather than replaces, human decision-making processes. In this setting, posts with privacy settings that may diverge from the intended one are flagged for the user, similar to Facebook's "people you may know" interface. Our predictive model achieved a 317% improvement in accuracy (precision-recall AUC) when compared to simple prediction rules such as limiting sharing for friends with low levels of interaction. The predictive power of a variety of features (including user features, post statistics, the post's content, and characteristics of the audience) show the importance of friend context in predicting the correct privacy setting. Crucially, we found that the most predictive features can be collected without human interaction.

Surprisingly, observable friendship dynamics like the frequency of interaction on Facebook or length of friendship alone are insufficient as predictors. The former was weakly correlated with privacy preferences, and the latter was not significantly correlated with privacy preferences at all. Participants often wanted to share with Facebook friends with whom they never visibly interacted, some of whom were close offline friends or family members.

While a few prior studies found that users need to retrospectively revisit Facebook privacy settings [4, 6], we take a holistic, user-centric approach to unpacking this problem within the context of a user's entire Facebook history, including the dynamics of changing sets of Facebook friends. We also take the first concrete steps toward building human-in-the-loop interfaces that use predictive models to identify posts whose privacy settings the user ought to revisit.

2 FACEBOOK PRIVACY SETTINGS

Facebook users control access to their posts by choosing privacy settings with the Audience Selector [22]. While the particular settings Facebook provides have changed substantially over the years, they have encompassed granting or denying access to both *individual users* and to *roles* (e.g., the user's Facebook friends, user-specified groups of friends, users tagged in a post). Just as in traditional role-based access control (*RBAC*), roles like 'friends' or 'users tagged in this post' describe sets that shift over time. Previously, permissions could be granted to a user's *networks* (e.g., University X). This option has since been removed. We focus on the following five settings that specify to whom Facebook content is accessible:

- **public** (previously "everyone"): anyone on the web [20]
- **friends+**: the user's Facebook friends plus the friends of some/all of those friends (e.g., friends of friends, friends plus anyone tagged) [19, 40]
- **friends**: the user's Facebook friends [40]
- **custom**: a user-specified subset of Facebook friends [19]
- **only me**: only the user [19]

In addition to changing the available options over time, Facebook has also varied the default, complicating longitudinal privacy management. In 2008, the default was *friends plus networks* [40]. The default was changed to *public* in 2010 [54] and *friends* in 2014 [50].

3 PROPOSED IMPLEMENTATION

This paper reports on a user study designed to build a longitudinal understanding of Facebook privacy attitudes and practices, as well as an investigation of how preferences correlate with various properties of posts, users, and settings. The ultimate goal of building this increased understanding of privacy settings over time is to build a human-in-the-loop retrospective privacy management system. In such a model, suggested privacy setting modifications would be presented to users through an interface that closely mirrors the "people you may know" feature on many social media sites.

With such an interface in mind, the objective we wish to maximize in this work is not pure accuracy, but rather a balance between accuracy and the importance of the suggested change. Regardless of the accuracy of such a prediction service, users must retain agency over important decisions like adding friends or revoking access to shared posts. An important implication is that while false negatives are certainly unwanted, the cost of such an incorrect suggestion is less catastrophic than in other security and privacy contexts, such as intrusion or spam detection. Furthermore, as this is a maintenance task, this suggestion interface can complement direct management tools like Facebook's "privacy checkup."

4 RELATED WORK

Broadly, privacy settings on social media can be considered a form of RBAC, which allows policies that specify permissions based on a user's role (e.g., "manager" or "contractor") [64, 67]. Access control policies can be complex, as documented in studies of system administrators [7, 8]. A rich literature has proposed many techniques for helping users accurately specify and audit access control policies. These techniques include matrix-style visualizations [62], rich queries of the authorization server [84], decision-support systems [11, 14], and human-in-the-loop iterative refinement of policies [36]. Researchers have also proposed alternate ways of expressing access control policies based on context [41], just-in-time requests [51], and semantic tags [38, 52].

Mismanagement of Facebook privacy settings can be caused by user misunderstandings [1, 15, 49], mismatches between the actual and expected dissemination of content [9, 12, 45], and overly complex user interfaces [33]. In 2011, Liu et al. surveyed 200 Facebook users, finding that 63% of posts were exposed to a larger audience than desired [47]. While users sometimes choose not to share content proactively [68] or delete content [2, 58], the mismanagement of privacy settings can cause embarrassment and regret [69, 76].

4.1 Longitudinal Privacy on Facebook

Use of a social media platform changes considerably over time. Backstrom et al. noted a significant turnover in a user's set of close Facebook friends [5], causing a "time collapse" in which temporal context is lost [13]. Privacy behaviors also change. Stutzman et al. found increased non-public content in Facebook profile attributes (e.g, date of birth) over time [73]. Users themselves also change [4].

We observe that, because of friend addition or deletion, the number of people included in these settings also implicitly changes over time. In RBAC parlance, the *friend* role is granted to, or revoked from, different users at different times. This change is automatic. A post made in 2009 and shared with *friends* might be visible to 150 users when created, but friend additions may cause it to be shared with 1,500 users in 2019 without any privacy setting changes.

These longitudinal changes in platforms, combined with users' life and relationship changes, necessitate retrospective management of privacy settings [59]. Prior work found that although access control settings in corporate environments rarely need to change [70], access control settings chosen long ago are frequently inaccurate in both social media [3, 4, 6] and cloud storage [37].

Two closely related studies have documented the need to revisit privacy settings for past posts. Through user studies leveraging the Facebook API, both Ayalon and Toch [4] and Bauer et al. [6] showed participants past posts. In the former study, participants answered questions about their likelihood to edit or hide the post. In the latter study, participants answered questions about their desired future audience for the post. These studies found that life events and the passage of time are weakly correlated with desired changes to a post's audience. The first part of our study partially revisits this work. However, we collect a far larger and richer set of features. We also explicitly show participants a given post's current privacy setting during the study and ask whether they would actually want to change it. We also use the full history of interactions between a user and each of their Facebook friends to further understand the longitudinal evolution of privacy settings. In contrast to the previous work, our work also aims to build predictive models for identifying posts with currently inaccurate privacy settings.

Facebook and similar platforms provide few options for retrospectively reevaluating privacy settings. In 2011 Facebook introduced a "limit past posts" feature that changes all posts shared beyond the user's Facebook friends to the friends-only setting [31]. The "privacy checkup" feature, introduced in 2014 [50], lets users examine and change their default privacy setting. While these tools can be effective, they unilaterally update sharing settings for large sets of posts or friends. We instead focus on finding specific posts whose privacy settings are likely to be inaccurate. Revisiting old posts is also facilitated by Facebook's "on this day" feature, which highlights posts from a given date in earlier years [34]. However, it neither provides a global view of aging posts nor offers assistance on retrospectively managing privacy.

4.2 Helping Users Choose Privacy Settings

Researchers have proposed a number of strategies to help users choose privacy settings. These techniques include audience-centric views of a post [44] and the ability to assign Facebook friends to custom groups (e.g., "band people") [35]. Variants of both have since been adopted by Facebook. Researchers have also suggested new visualizations of privacy settings [16, 53] and automated "nudges" highlighting a post's potential audience and impact [75, 78].

Some researchers have also proposed using machine learning to predict a post's initial privacy setting. For example, Fang and LeFevre use active learning and friend clustering to predict fine-grained privacy settings [23]. Others have built predictive models for computing inter-user tie strength [28], user-level privacy scores [46], privacy risk [83], and the privacy similarity between users [27]. More recently, Fiesler et al. built a logistic regression model to predict whether or not a post should be public [24]. Supervised learning has also been used to understand private information disclosure attacks in online social networks, specifically for sensitive attribute inference [26, 29, 39, 43, 82], sensitive relationship inference [10, 81], and identity matching across platforms [80].

While these efforts focus on helping users choose the initial privacy setting for a post at the time it is posted, we instead focus on helping users identify posts where this initial privacy setting no longer matches the currently desired setting. We build on these prior models by incorporating features they found to be predictive, adding other features, and testing different model architectures. A frequently proposed, and sometimes implemented, idea is to instead let users set an "expiration date" when making posts [3] or to otherwise set a time limit on information sharing. For example, Snapchat messages disappear after a matter of seconds [56], Instagram stories disappear after 24 hours [32], and the visibility of WeChat moments can be restricted to three days, six months, or forever [77]. However, prior work has found that users' predictions about future changes in the visibility of their posts frequently do not match their later preferences when revisiting those posts [6].

5 USER STUDY METHODOLOGY

To understand Facebook users' longitudinal privacy attitudes and practices, document the degree to which retrospective reevaluation of privacy settings is needed, and collect the data needed to build and test predictive models for helping users do so, we conducted a user study. We investigate these questions using data collected in two surveys alongside an anonymized version of each participant's full Facebook timeline and activity log (collected with consent).

5.1 Recruitment and Survey 1

We recruited participants from Amazon's Mechanical Turk who were located in North America, 18+ years old, and had a 95%+ approval rating. We screened participants using their account data to verify that they met our inclusion criteria. These criteria, also listed on the study advertisement, were: (i) had a Facebook account for at least 2 years and (ii) made at least 10 posts in the last year.

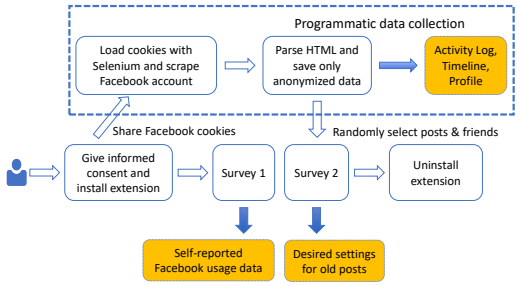


Figure 1: Our protocol and process of data collection.

Category	Time of post	Now	Median # friends
X-Low	Not Facebook friends	No visible interaction	23
X-High	Not Facebook friends	Frequent interaction	4
Low-Low	No visible interaction	No visible interaction	35
Low-High	No visible interaction	Frequent interaction	6
High-Low	Frequent interaction	No visible interaction	2
High-High	Frequent interaction	Frequent interaction	4

Table 1: Categories for selecting specific people in Survey 2 based on whether they were Facebook friends with the participant at the time of the post, as well as how frequently they interacted (likes/comments/tags) with the participant in the year prior to the post and the year prior to the survey.

These ensured the accounts were sufficiently well-used for us to investigate retrospective privacy management.

Participants volunteered for the study, installed our browser extension (see below), and then took Survey 1. This survey asked about the participant's overall Facebook usage, their use of Facebook's privacy features, and their demographics. Participants were compensated $10 for each survey. Appendices A–B contain the survey instruments.

5.2 Ethical Collection of Facebook Data

Both measuring longitudinal behaviors and building predictive models necessitate the collection of participants' Facebook data. Our goals for this data collection were to collect data in the most privacy-preserving way possible and to obtain very explicit and fully informed consent from participants for doing so. We first considered having participants use Facebook's "download your data" feature and uploading this full set to our servers. While full consent would be possible, we would inadvertently be writing private data (including Facebook messages and ad clicks) to disk, so we rejected this option. Alternatively, we considered partnering with Facebook under their blanket TOS agreement for data collection. We rejected this option because prior research using this approach, including the widely discussed 2014 Social Contagion study, demonstrated barriers for giving meaningful consent [66].

We thus decided to design a protocol leveraging a browser extension that elicits meaningful informed consent from participants and collects data in a privacy-preserving way. This protocol was approved by our IRB. Figure 1 summarizes our data-collection infrastructure. This infrastructure enables us to collect, with the participant's permission, anonymized versions of their full Facebook timeline (posts they previously made), as well as their Facebook activity log. The former enables us to programmatically analyze their prior posts, as well as those posts' privacy settings and metadata (e.g., likes, comments). The latter enables us to analyze the temporal evolution of their set of Facebook friends and similar events.

Because of the nature of the data we were collecting, we did not consider a standard consent form sufficient. Thus, after participants agreed to our standard consent form, we provided a separate page detailing the data we would and would not collect from their Facebook account, including visual examples of our anonymization procedures (described below) and an overview of our technical approach. If the participant wished to continue, they then downloaded a browser extension we designed. This extension shared their Facebook session cookie with a server at our institution. Our servers used the Selenium browser automation tool [65] to download the relevant parts of a participant's timeline and activity log.

Programmatic data collection: We performed all data collection programmatically. Researchers never viewed the raw HTML of any participant's account. In Survey 2, we embedded links to the Facebook URLs of posts and profiles rather than saving or serving any potentially sensitive content on our servers.

Anonymization: We only stored anonymized versions of the data by using one-way hashes for any unique identifiers (e.g., numeric Facebook IDs, names of any Facebook users) that could be considered personally identifiable information (PII). We also did not analyze or store any photos included in posts. We performed this anonymization procedure before writing data to disk. Despite our best efforts, we acknowledge that some collected data (e.g., the content of a post) might still contain PII (e.g., a nickname) that is very hard to detect automatically. Our anonymization strategy is similar to what Facebook themselves adopted via App-Scoped IDs in their API [17]. We never tried to deanonymize any account.

Targeted data collection: We did not collect data from any Facebook page that is not part of a participant's account. Specifically, we did not collect information posted by participant's friends. We collected aggregate data on likes and comments made on the participant's timeline, as well as the participant's own likes and comments. In contrast to earlier studies [28, 57], we chose not to collect potentially useful data on the structure of the social graph from participants' friends who had not volunteered to participate.

5.3 Survey 2

Survey 2 contained two parts. In Part 1, we embedded links to five randomly selected posts from the participant's timeline. We showed the post's current privacy setting and asked if the participant had ever changed (or considered changing) that setting. We also asked participants whether they wanted to keep that privacy setting or choose a different one moving forward, and why. Different from earlier studies [4, 6], we chose to remind participants of their current privacy setting for each post and make "keep this setting" the first option. While this could prime users to keep their current privacy setting, keeping the current setting minimizes friction.

In Part 2 of Survey 2, we revisited those same five posts, this time showing the participant six specific Facebook friends who could currently see each post. We asked whether or not the participant preferred to continue sharing that post with that person moving forward, or whether they did not care (so as to differentiate strong preferences from indifference or a default preference).

Based on our observation that the "friends only" privacy setting changes meaning as a user adds friends and our hypothesis that participants might want to stop sharing content with Facebook friends with whom they never interact, we used stratified sampling to select the six Facebook friends. For each post, one friend was randomly selected from each of the six categories enumerated in Table 1, which capture temporal changes in how the users interact, as well as whether the users were friends at the time of the post.[1] Sampling based on visible interaction was inspired by Gilbert et al., who used similar interaction features to measure tie strength [28]. We calculated the level of interaction as the sum of: (i) the number of words exchanged via timeline posts and comments; (ii) the number of intimate [74] words exchanged; (iii) the number of posts on each other's timelines; and (iv) the number of likes and reactions the participant and their friend gave each other's posts. We divided friends into high (top 10% of friends) and low interaction (no visible interaction) buckets and exclude those who fall in neither. For each post, we computed visible interaction within two time spans: the year before our study, and the year before the post was made. For each post, we thus randomly sampled one friend per category. At the conclusion of Survey 2, we instructed participants to uninstall the plugin and log out of Facebook to invalidate their session cookie.

5.4 Data Analysis

As detailed when we present results, we performed statistical testing to investigate our targeted quantitative hypotheses. We also built and evaluated statistical models using standard evaluation measures like accuracy, precision, and recall.

For consistency, two researchers independently coded free-text responses using a shared codebook. Across questions, Cohen's κ (inter-rater agreement [42]) ranged from 0.7 to 1, indicating substantial to perfect agreement. The coders met to resolve disagreements and choose a final code for each response.

5.5 Limitations

A core limitation is that we used a convenience sample of North American MTurk workers, and this sample consisted of only 78 participants. Nevertheless, our sample still contained participants with wide variations in account age and daily usage. Furthermore, our results likely underestimated privacy needs as highly privacy-sensitive individuals would be unlikely to participate in our study. However, even our participants wanted to restrict the visibility of 13.9% of posts they saw. In our study, we recruited English-speaking US Facebook users to enable comparisons to prior work. Thus, our results may not generalize to users from other languages or countries. As we only consider visible interactions on Facebook, we inevitably miss offline interactions. Our goal, however, is building predictive models that leverage only online data.

6 RETROSPECTIVE PREFERENCES

Here, we characterize participants' retrospective access control preferences for their old Facebook posts. Broadly, our results reinforce the need for automated assistance in reviewing access control (privacy) settings for aging Facebook posts.

[1]We did not collect deleted Facebook friendships. Hence there are no Low-X, High-X categories in Table 1.

Characteristic	Total	Min.	Median	Max.
Account Age (Years)	-	3	10	13
Friends	-	12	224	3,625
Timeline posts	253,122	87	1,840	15,470
Non-timeline activities	1,738,303	1,509	20,263	60,184

Table 2: Overview of participants' Facebook accounts.

6.1 Participants' Demographics

A total of 101 participants installed our plugin and completed Survey 1. However, 13 participants did not meet our stated inclusion criteria (e.g., based on the age of their account), so we did not invite them to participate in Survey 2. Of the remaining 88 participants, 78 completed Survey 2, and those are the responses we analyze. For these 78 participants, we collected preferred privacy settings for 390 posts and 2,340 friend-post pairs (see Section 5). The posts for which participants answered our questions had a median age of 2.6 years (minimum 9.9 days, maximum 9.9 years).

Basic demographics: Among participants, 69% identified as female, and the rest as male. This skews more female than Facebook overall (52% female in 2018 [71]). A plurality (46%) of participants were in the 25–34 age range, and the overall age distribution is consistent with Facebook users overall in 2018 [72]. 87% of participants identified as white and 9% as black. 18% of participants held a degree or job in computer science or a similar field.

Facebook usage: Among participants, 89.7% reported daily Facebook use, with median usage of 1 hour per day. We summarize participants' Facebook accounts in Table 2. While we did not specifically attempt to recruit users with especially long-lived accounts, 55 of our 78 participants' accounts were at least 10 years old. In aggregate, participants made a total of 253,122 posts on their own Facebook timeline, with the median participant making 1,840 posts. Their activity logs showed participants also performed significant non-post activity, such as liking other posts or watching videos.

6.2 Temporal Changes in Facebook Usage

In Survey 1, we asked participants to report the primary topics of their Facebook posts at three points in time: one year after they initially created their account; at the time of the survey; and halfway between the two. At each point, participants reported posting about their personal lives, including humorous content and updates about their family. Only 5 participants (6.4%) said posting updates about their personal lives was their primary Facebook activity at the time of the study, whereas 15 participants (19.2%) said it was their primary Facebook activity at the midpoint of their account lifetime.

We also asked participants which year they thought their Facebook usage peaked, and how their current usage compares with their peak usage. On average, participants reported their Facebook usage (amount of time spent on Facebook) peaked 5.6 years ago (σ = 3.4, median = 5). Two-thirds of participants reported that they currently spend less time on Facebook than they did during the peak year. This evolution in Facebook usage further motivates the need for retrospective privacy-management tools.

Increasing audiences: For the 55 participants whose Facebook accounts were at least a decade old, we analyzed how their set of Facebook friends changed over time. Figure 2 shows the percentage of participants' current friends added over the past years (using

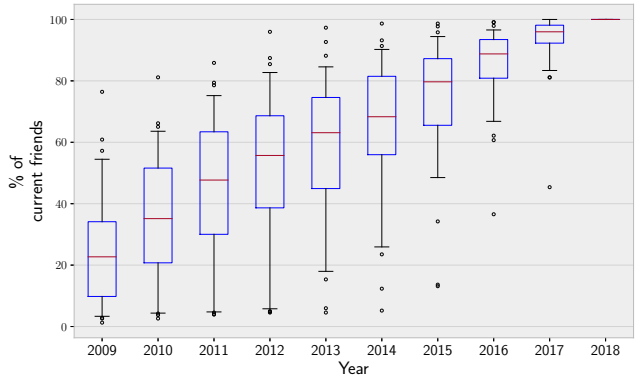


Figure 2: Percentage of 2018 friends who were friends in previous years for the 55 accounts at least 10 years old.

friend-addition timestamps in the activity log). The boxes represent quartiles. We observe a very substantial increase in the audience implicitly included in the most common "friends only" privacy setting. Compared to 2018, the median participant had under half as many friends in 2012, and under one-quarter as many in 2009. Figure 13 in the appendix is an analogous graph for all 78 participants, showing an even more pronounced trend.

Offline Events Affect Facebook Usage: A Facebook user's life changes impact which privacy settings they desire [4, 6]. 56.4% of our participants mentioned that life events affected their sharing decisions. Specifically, they reported that their sharing on Facebook was affected by personal (30 participants), professional (10), and global (18) events. Relationship changes (10) and childbirth (7) were the most frequently mentioned personal changes. Career changes (3) and issues with a coworker (3) were the most frequently mentioned professional changes. Finally, elections (6) and news about data breaches (6) were the most frequently mentioned global events motivating changes. Some of these events led to fewer personal posts on Facebook (reported by 13 of the 30 participants with personal changes and 3 of the 10 with professional changes).

Usage of Privacy Features: We also investigated usage of Facebook's own retrospective privacy features. 59% of participants reported that they had seen Facebook's "privacy checkup" tool when we showed them a picture of it, and 44.9% reported that they had used the privacy checkup. These high percentages are consistent with a 2018 Reuters/Ipsos survey that reported 74% of U.S. Facebook users were aware of their current Facebook privacy settings [63]. Similarly, 53.8% of our participants reported seeing the "limit past audience" feature. However, only 19.2% recalled using it.

6.3 Privacy Settings Over Time

Figure 3 shows the distribution of privacy settings for participants' posts made each year from 2009 to 2018. The x-axis labels indicate the total number of posts made each year. The result is similar if we include shorter-lived accounts (Figure 14 in the appendix).

We found that *friends* was by far the most used privacy setting, even for posts made pre-2011 when posts were *public* by default. This finding appears to contradict earlier work from Liu et al. that found that the *public* setting is most heavily used on Facebook [47]. Note that Liu et al. also used Mechanical Turk, surveying 200 users. If both our and Liu et al.'s samples are sufficiently representative, our observed distribution of privacy settings on pre-2011 posts suggests a recent (and significant) restriction of the visibility of old posts. 2011 also saw the introduction of the "limit past posts" feature, which our participants might have used to restrict their old posts' visibility. While only 19.2% of our participants reported that they remembered using that tool, they could have forgotten having done so, or they could have restricted posts manually. Such a significant change in privacy settings constitutes a major incident regarding retrospective privacy. Unfortunately, "limit past posts" can only restrict widely shared (e.g., "public") posts to "friends only." It is an incredibly blunt tool and cannot capture subtle, and sometimes important, retrospective privacy decisions.

6.4 Desired Privacy Settings

Table 3 presents the results of asking whether participants wish to change the privacy settings for five randomly selected Facebook posts. We exclude one post where the participant preferred not to answer about his desired setting. We make two observations from this table. First, while a majority of existing privacy settings for 290 old posts (74.5%) do not require changes, 65.3% of participants wanted to change at least one post's privacy setting. The gray-colored cells of Table 3 indicate posts where participants did not want to change their current privacy settings. Second, we found that preferred changes in settings are roughly split between increasing and decreasing the audience size. Earlier work [6] reported similar results. Interestingly, even for two posts currently shared with custom settings, the participants wanted to share them with different custom settings containing smaller audiences.

The red and blue shaded regions in Table 3 indicate a decreased or increased audience, respectively. When we asked participants why they want to change their privacy setting, the most common reasons were that the post was not appropriate (18 posts), it was irrelevant (16), and they did not care who was able to see that specific post (14). When participants wanted to increase the post audience, they mentioned that it was because it contained public information or a general message that they would like more individuals to see.

When participants were asked how important it was to change each post's privacy setting, 65 changes were of only slight or no importance, and 34 changes were of extreme or moderate importance. If participants felt that changing the privacy setting of a post was extremely important, they often attributed this to the post being inappropriate or containing private information. If participants felt that it would be very or moderately important to change privacy settings, appropriateness was still an important rationale. For instance, one participant commented, *"It was a trip with my ex, I doubt my fiancee wants to see that."*

6.5 Reasons for Retrospective Changes

Participants indicated that they wish to change the privacy settings of 25.5% of their old posts. In this section, we further investigate how they came to these conclusions.

6.5.1 Retrospectively Browsing Old Facebook Posts. To gain insight into abstract concerns regarding others browsing old posts, we asked participants about their perception of, and participation in,

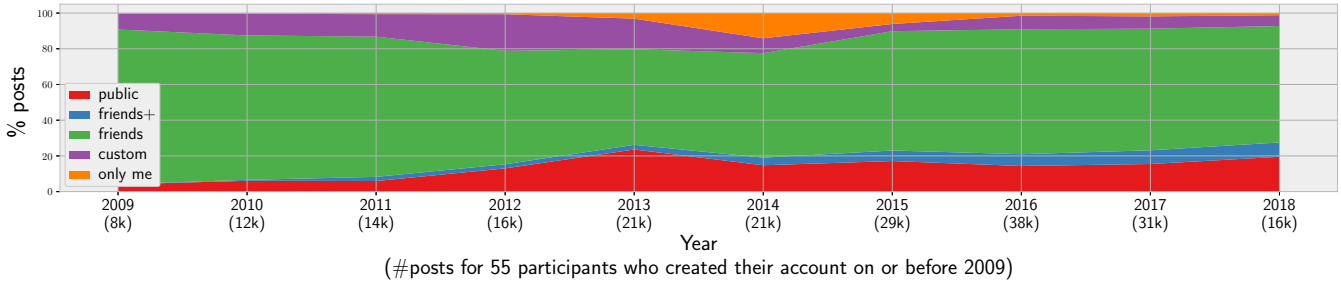


Figure 3: Among the 55 participants with decade-old accounts, the percentage of their posts in each year with each privacy setting. Each participant's set of posts occupies an equal amount of y-axis space. The majority for all years is "friends only."

	Desired setting							Audience			
Current setting	**Public**	**Friends+**	**Friends**	**Custom**	**Only Me**	**Custom (Decreased)**	**Delete**	**Total**	**Increased**	**Same**	**Decreased**
Public	58	-	3	-	-	-	1	62	-	58	4
Friends+	3	27	3	-	-	-	-	33	3	27	3
Friends	21	4	177	3	5	-	31	241	25	177	39
Custom	6	2	9	19	1	2	4	43	17	19	7
Only Me	-	-	-	-	9	-	1	10	-	9	1
Total	88	33	192	22	15	2	37	389	45	290	54

Table 3: Comparison of current and desired privacy settings for the 389 posts in Survey 2, excluding the one for which the participant preferred not to answer. For two posts with a custom setting, participants chose a new custom setting with a smaller audience. Gray denotes keeping the same setting, red denotes a smaller audience, and blue denotes a larger audience.

browsing old posts on Facebook. The results reported in this section are based on Survey 1 data. This data is not grounded in particular posts, but rather participants' general perceptions about friends/themselves browsing old posts. First, our participants expect this browsing to happen: 67 participants (85.9%) believe that some or most of their friends will browse their profile and check old posts. And while only 9 participants (11.5%) reported that they would feel uncomfortable if their friends browsed their one-year-old posts, 22 participants (28.2%) reported feeling uncomfortable if their friends were to browse their three-year-old posts. In contrast, 43 participants (55.1%) reported checking their friends' one-year-old posts, and 18 (23.1%) reported checking their friends' three-year-old posts. Moreover, a small number of our participants self-reported arguably invasive behaviors in browsing friends' profiles, including checking relationship history (1 participant), stalking (3), fact-checking (1), and digging up family information (2). These intentions certainly motivate retrospective control of post privacy for even the most slightly privacy-conscious Facebook user.

6.5.2 Effectiveness of Existing Mechanisms. Finally, we evaluated the effectiveness of Facebook's current privacy-management mechanisms by checking for a correlation between whether a post's existing and desired privacy settings differ and whether the corresponding participant had used an existing privacy-preserving mechanism. We used the χ^2 test [60] or Fisher's exact test [25], depending on the amount of data available for the individual test. We did not find any significant correlations between the frequency with which participants wanted to change the privacy settings on old posts and their use of various privacy-preserving mechanisms (removing a friend, changing the audience of a past post, and using the "privacy checkup" [21] or "limit past posts" [18] features).

This lack of a significant correlation suggests that the currently available tools are insufficient. If they were meeting users' needs for retrospective privacy management, we would expect to see less need for changes when users were already actively managing their privacy. This analysis is potentially confounded by participants not being able to remember seeing or using these tools, especially in light of the observation that many participants had likely used these tools to retrospectively change privacy settings (see Section 6.3). Even so, the lack of clear evidence that the current tools are sufficient is motivation for creating new tools that can support users' clear need to manage these privacy settings retrospectively.

7 PRIVACY PREFERENCES' CORRELATIONS

Our study is unique among research on social media privacy in its combination of temporal reach (with the median participant's account being 10 years old) and the fine-grained nature of the data, contextualizing privacy preferences within the participant's full timeline and evolution of their set of Facebook friends. Here, we explore correlations between retrospective privacy preferences for continuing to share given posts with given Facebook friends based on both how frequently the two users visibly interact on Facebook and when that relationship was added on Facebook.

7.1 Visible Interaction

Figure 4 presents participants' preferences for sharing particular posts with particular Facebook friends based on their relationship (friends or not at the time of the post) and evolution in visible Facebook interaction with that friend. The top panel presents preferences for each of our six friendship categories. The middle panel considers only interaction over the year before our user study, while

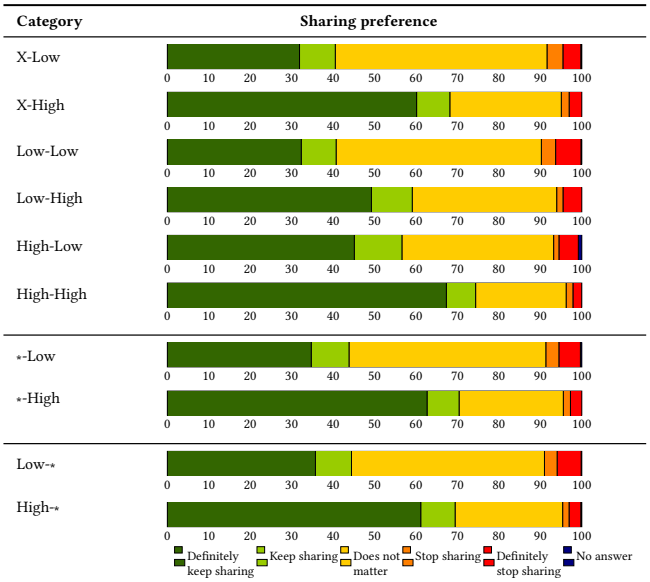


Figure 4: Sharing preferences by friend category (cf. Table 1). The first part of a name refers to the year before a post was made, while the second refers to the year before the study.

the bottom panel considers only interaction in the year before the post was made.

Figure 4 shows that participants were more likely to "definitely keep sharing" posts with friends with whom they had a high degree of recent interaction than those with whom they had not recently interacted (Mann-Whitney U, $p < 0.001$). They reported definitely wanting to keep sharing posts with 62.7% of friends with whom they frequently interacted in the year preceding our study. While recent interaction is indicative of a desire to definitely keep sharing, the inverse is not true. Participants also wanted to "definitely keep sharing" posts with 34.7% of the friends with whom they had no visible interaction on Facebook in the year before the study.

We used our qualitative data to better understand the desire to share posts in spite of no recent interaction. Often, the participant identified the Facebook friend as a family member or close friend, which implies that the level of visible interaction on Facebook is an imperfect measure of real-world closeness. Additionally, participants sometimes anticipated that the content of a post would be interesting to the friend. This mental model of friends' interests is not reflected in our interaction data. Two less frequent reasons for sharing with friends in spite of no recent interaction are content-centric. For instance, participants wanted to keep sharing posts containing informative or humorous content with their friends regardless of visible interaction. We note very similar reasons when investigating prediction inaccuracies in Section 8.5.

Similarly, participants were more likely to definitely want to stop sharing with friends they had not recently interacted with (8.3% of the time) than those they interacted frequently with (3.5% of the time) (Mann-Whitney U, $p = 0.001$). The similarity of these numbers underscores that while interaction is correlated with sharing preferences, it is not sufficient on its own for prediction, as explored in future sections.

7.2 Correlation with Time of Friendship

Recall that the median participant had twice as many Facebook friends in 2018 as in 2012, substantially changing the meaning of a "friends only" privacy setting. Surprisingly, we did not observe significant differences in whether a participant wanted to share a given post with a given friend based on whether or not they were Facebook friends at the time the post was made. The "X-*" plots in Figure 4 depict this phenomenon. In other words, the time of Facebook friendship is not only insufficient for retrospectively predicting whether a post should be shared with a given friend, it does not even seem to be correlated. While we had initially hypothesized that participants might not want to share past content with friends they make in the future, our results do not support this hypothesis. Instead, participants appear to be adding new Facebook friends with the intention that these new friends can access past content.

8 PREDICTING PREFERENCES

Our ultimate goal is to enable users to efficiently maintain correct privacy settings on years or decades of social media posts. The sheer number of friend-post combinations for even light social media users necessitates automated assistance for this task. To support the use of machine learning models in such a subjective and important setting, we leverage insights regarding preferences from Sections 6 and 7, designing models intended for use within the privacy domain and the user assistance scenario.

8.1 Prediction Task

For the prediction task, our dataset consists of tuples $(\mathbf{X}_i, Y_i)$, where $\mathbf{X}_i$ is the feature vector and Y_i is the desired audience change for post i. We formulate the problem as a binary classification task where $Y_i = 1$ corresponds to *limit sharing* and $Y_i = -1$ corresponds to *do not limit sharing*. Our task is binary classification, since our current focus is to help users find posts they wish to limit sharing, based on a human-in-the-loop system, rather than building a fully automatic post manager. By mapping our problem to binary classification, we can get a better separation on posts users specifically wanted to *limit sharing* compared to *do not limit sharing*. After prediction, we can sort posts by their likelihood of *limit sharing* to show users posts in the predicted priority order.

The feature vector $\mathbf{X_i}$ includes variables capturing the survey responses, including some user information, post statistics, content, and audience. From the survey features, we have the age of the account and the age of the participant as user information. We include the survey responses either as one-hot encoding or binary indicators for multiple-choice responses. Our post statistics features are the following: the number of likes and comments, the content type (e.g., text, link, image), whether another user is tagged, if comments were edited, if the audience was changed earlier, the age of the post, and the current privacy setting. We extract content-level features from the text of posts through established NLP feature extractors: Google News Word2Vec embedddings [55], Linguistic Inquiry & Word Count (LIWC) categories [74], Google's content classification categories, and Google's sentiment scores (i.e., positive or negative sentiment) [30]. Our audience features include friend-specific features: days since first and last communication, reaction counts, wall words exchanged, and how many wall posts the user

initiated to a friend. We include more details on features in Table 6 in the appendix.

To perform binary classification, we compare several established supervised learning algorithms: Decision Trees (*DT*), Logistic Regression (*LR*), Support Vector Machines (*SVM*), Random Forests (*RF*) using scikit-learn [61], and XGBoost (*XGB*) [79]. We also include Deep Neural Networks (*DNN*) using scikit-learn and the Adam optimizer, although DNNs tend not to learn well from small datasets like ours. For our DNN, we used 3 hidden layers with 100, 50, and 20 nodes with RELU activations and a softmax activation for the output layer. We report results only on the best performing classifiers, while leaving results for other classifiers in the appendix (Figures 10, 11, and 12). In the absence of any preexisting classifier, we propose two baseline models. The first is a random classifier (*Random*), where we randomly show posts to users. The random classifier is used when there is no information for predicting if a post will be selected for *limit sharing*. We also considered a more reasonable straw man baseline (*Interaction*) that does not require machine learning, but only considers the level of interaction between the user and their friend. This baseline predicts *limit sharing* for friends with low levels of interaction. We chose these baseline classifiers because, to the best of our knowledge, no prior work has attempted to predict posts and friend-post pairs for which to retrospectively limit sharing.

8.2 Dataset Description

We consider two datasets for predicting privacy preferences. In the post dataset, we aim to predict whether a user should decrease the audience of a post. In the friend-post dataset, we aim to predict whether a user should remove a specific friend's access to the post. For both datasets, we focus on the binary classification task of predicting whether or not a user wishes to *limit sharing*.

Post Dataset. In the **post dataset** there are 389 posts for which users specified labels. There are three labels in the dataset: *less*, *same*, and *more* audience. Since we focus on finding posts for which the user wishes to decrease the audience, we treat *less* audience as *limit sharing* and the other two as *do not limit sharing*. We have the following label distribution: 13.9% for *less*, 74.5% for *same*, and 11.6% for *more* audience. For binary classification, we have: 13.9% for *limit sharing* and 86.1% for *do not limit sharing*.

Friend-Post Dataset. The **friend-post dataset** contains the same posts as the **post dataset**. However, participants specified audience-change labels for specific friends (up to 6 friends per post). This dataset contains 2,336 total labels, after removing friend-post pairs where no answer was given. This dataset contains 3 possible decisions for privacy preference: *stop sharing*, *doesn't matter*, and *keep sharing*. We map this to a binary classification task where *stop sharing* corresponds to *limit sharing*, and the other two correspond to *do not limit sharing*. For friend-post pairs, we have the following label distribution: 6.4% for *stop sharing*, 36.4% for *doesn't matter*, and 57.2% for *keep sharing*. For binary classification, we have: 6.4% for *limit sharing* and 93.6% for *do not limit sharing*.

Both datasets are highly skewed toward *do not limit sharing*. This can highly bias our results towards predicting *do not limit sharing* for every post. We counteract this issue by focusing on the binary classification task, since we wish to discriminate posts that are *limit sharing* from all other posts.

8.3 Experimental Setup

In our experiments, we perform 5-fold cross validation and report averaged results across 5 testing folds. Since, we are focusing on finding posts where the user may wish to decrease the audience size, we order examples in the test data by the probability of being $Y_i = 1$ (*limit sharing*) and assess their precision and recall. This is a typical evaluation setup for binary classification where one label (*limit sharing*) is more important than the other (*do not limit sharing*). Since we can vary the number of posts that we predict as *limit sharing*, we report on precision@k, the precision after predicting the top k results as positive. Each value of k is considered a potential cutoff, where all examples ranked greater than or equal to k are classified as positive and the rest are classified as negative. We compute precision as $TP/(TP + FP)$, where TP is the number of true positive examples (actual label positive, predicted label positive) and FP is the number of false positive examples (actual label negative, predicted label positive). Thus, precision@k is the proportion of correctly classified positive examples for all examples above the cutoff k. In other words, the precision@k is the binary precision when only considering the top k examples. Precision@k curves allow us to see how accurately we are predicting our desired label after showing to users the most likely posts for decreasing the size of the audience. We also compute recall as $TP/(TP + FN)$, where FN are the false negative examples (actual label positive, predicted label negative). We report precision-recall curves to show the tradeoff between showing a larger number of posts that need users' attention and how accurately we can uncover such posts.

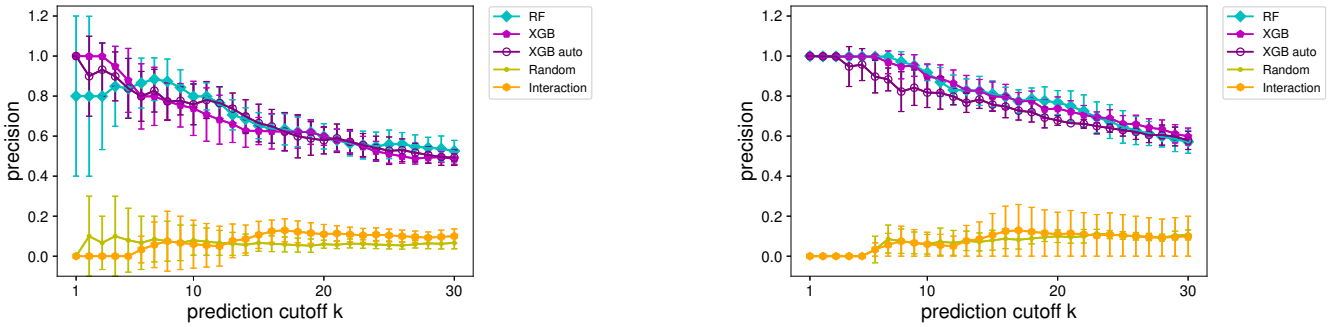
Ordering examples by the probability of correctness also maps well onto an implementation that mimics the "people you may know" feature employed by Facebook and other social networks. Prioritizing the suggestions that are most likely to be correct maximizes the utility of the tool in an environment constrained by user attention. Furthermore, since it is unlikely that a user will be willing to spend the time to go over all suggestions, our intention is to minimize the number of false predictions rather than ensure that all posts needing correction are (eventually) suggested.

8.4 Results

We present the precision@k and precision-recall curves averaged over the five folds. We also analyze the features for predicting friend-post pair privacy settings.

8.4.1 Friend-Post Dataset Prediction. We study whether it is possible to predict if a user wants to *limit sharing* for a post with a specific friend. Thus, we include features about the inferred relationship between the user and the friend in addition to other features.

Figure 5a shows the precision@k curves for predicting privacy preferences in the **friend-post dataset**. Here, the ensemble classifiers Random Forest and XGBoost give the best precision, with XGBoost performing better for very low K. Since the underlying distribution of *limit sharing* for this dataset is 6.4%, a cutoff at that percentage would be reasonable in a deployed system. This corresponds to predicting the top 30 results per test fold where the precision@30 is 0.519 for Random Forest. Additionally, we include



(a) Average Precision@k on full test set

(b) Average Precision@k on test without "doesn't matter"

Figure 5: Precision@k curves for the friend-post dataset, comparing the original dataset and removing the label *doesn't matter* from the testing folds, using the best classifiers (XGBoost and Random Forest). The preference for doesn't matter is most of the interference for the precision@k curves.

another precision@k curve for XGBoost (*XGBoost auto*), where we remove features that are not automatically collected, such as survey responses. We see that even after removing these features, we can get very close precision@k curves to XGBoost on the full set of features. This shows promise in building a system, where we only need to know friend-post pair sharing preferences so that we can get more labels. Features can be collected automatically.

We analyze whether *doesn't matter* decisions contribute to most of the false positives in top positions. Figure 5b shows that after removing those examples from the test set, the precision becomes higher for all k and stays 1.0 for more top examples (6 vs. 3), compared to Figure 5a. This result implies that many posts for which users do not care to limit sharing appear near the top, which are more tolerable false positives than posts where the user actually does not want to limit the audience. Note that this figure is only for explanation purposes, as a priori knowledge of the *doesn't matter* class would not be possible in the real world. Thus, for performance purposes, Figure 5a presents the realistic evaluation. We further analyze false positives in Section 8.5.

To understand the tradeoff between false positives and false negatives in prediction, we perform precision-recall analysis. Figure 6 shows the precision-recall curve for friend-post pair predictions. For example, if we show the first 3 examples to users, we achieve 1.0 precision, which means all 3 examples are correctly labeled *limit sharing*. However, very low recall shows that we missed many posts for which users wish to limit sharing. If we set the cutoff to match the distribution of *limit sharing* (i.e., $k = 30$), then both the precision and recall are 0.49. If one were to compare this approach to a heuristic of suggesting posts to reevaluate based on a low level of interaction, the precision-recall area under the curve (PR AUC) is 0.118. Contrasted with XGBoost's 0.493 AUC value, this represents a 317% improvement over using the level of interaction with friends to predict sharing reevaluation.

While these accuracy and precision numbers would be unreasonable to deploy in a fully automated system, our intended deployment for this task is part of a human-in-the-loop system (see Section 2). Thus, we seek to achieve a balance of precision and accuracy, and

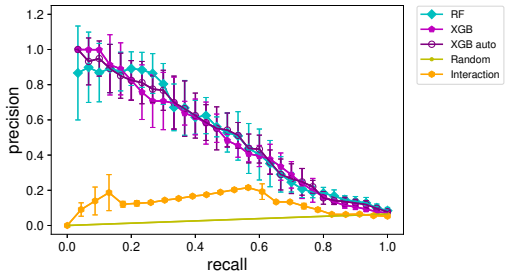


Figure 6: Precision-Recall curve for the friend-post dataset.

incorrect suggestions incur only a light time cost on users. Furthermore, deployment of such a system with satisfactory accuracy and precision would enable the collection of further user decisions to refine the performance of the classifier and suggestion mechanism.

Beyond simply achieving good performance, we also investigate which features are most predictive of the *limit sharing* decision. Table 4 lists the top 10 most important features according to XGBoost. From this list, we see that there is a mix of audience features (days since first and last communication, number of wall words exchanged, reaction counts), post statistics (age of the post, number of likes and comments of the post, whether the audience has previously been changed), and survey or user features (age of the account, user's number of friends, if the user had a personal life change). One notable result is that 9 out of 10 of these features can be collected without user interaction, while the other feature (if the user had a personal life change since the post) may require asking the user explicitly. Although not displayed here, some Word2Vec components and content classification categories were important, specifically in the top-20 features, while LIWC features and sentiment analysis did not appear to be highly important.

Next, we explored the effect of audience (or friend) context in the prediction. Figure 7 compares the precision@k curves when using all features, excluding friendship features, and relying *only* on friendship features using XGBoost. This suggests that while friendship context alone is insufficient, friendship features do play an important role in predicting friend-post pair privacy preferences.

Friend: Days since first communication with friend
Post: Age of the post
User: Number of friends
User: Age of the account
Friend: Days since last communication with friend
Post: Number of likes and comments on the post
Friend: Number of wall words exchanged from friend to user
User: If the user had a personal life change since the post
Post: If the audience of the post had changed previously
Friend: Reaction counts from the friend to the user

Table 4: Top 10 important features identified by XGBoost, sorted in descending of importance.

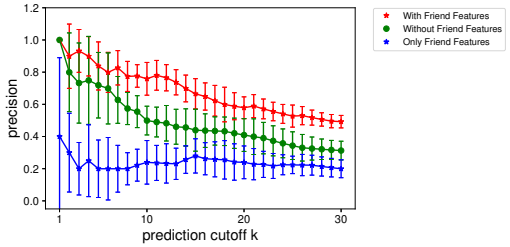


Figure 7: Comparing precision@k curves using friend features, no friend features, and only friend features for XGBoost. Post features are better than friend features individually, but combining them gives the best result.

We perform additional analysis on the neutral label *doesn't matter* because it was a large proportion of the friend-post pair dataset (36.4%). We do so by considering different variations for training our model, without changing anything in the testing data, with XGBoost as the classifier. Since the response *doesn't matter* is ambiguous, we consider treating it as different labels to see how the precision@k curves vary. We vary the training setup in four ways: (1) **original**: keep *doesn't matter* as the *do not limit sharing* category, which is our original setup; (2) **dm → limit**: treat *doesn't matter* as the decision to *limit sharing*, training with the original limit audience labels; (3) **dm → class**: treat *doesn't matter* as a separate class, transforming our problem into a three-class classification problem; (4) **dm → removed**: remove *doesn't matter* labels from the training process. In order to allow fair comparison across training setups, we use the exact same test data for all training setups and treat *doesn't matter* as *do not limit sharing* in the test data. For evaluation, we order test examples based on the probability of being *limit sharing*. Figure 8 shows that our original setup overall performs the best, especially for the top examples, while treating *doesn't matter* as its own class in training is a close second. This result is intuitive since we wish to identify posts to limit sharing, and separating them clearly from other examples during training will result in better classification. When we remove the label for *doesn't matter*, we get some decrease in precision. When treating *doesn't matter* as *limit sharing*, the precision@k drops significantly. The reason is that the classifier learns over two different types of labels for *limit audience*, which interferes with predicting the positive class during testing.

8.4.2 Post Dataset Prediction. Next, we study whether it is possible to predict if a user would want to *limit sharing* of a post entirely, rather than for specific friends. Figure 9 shows the precision@k curves for individual post prediction, using all classifiers. In this dataset, Logistic Regression performs the best. In Figure 9, the

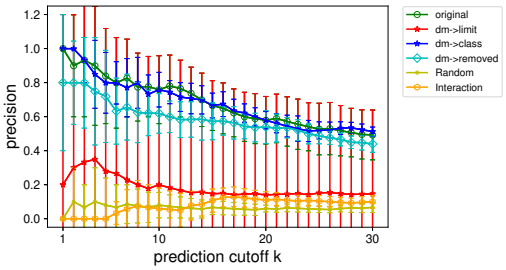


Figure 8: Comparing different methods for handling *doesn't matter* responses during training with XGBoost.

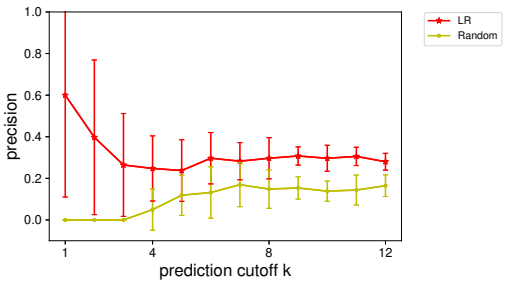


Figure 9: Average precision@k for post dataset.

precision is relatively low even for top results (low k). Since the underlying distribution of *limit sharing* for this dataset is 13.9%, a cutoff at that percentage would be reasonable in a deployed system. This corresponds to predicting the top 11 results per test fold where the precision is 0.288. The best classifier for this task is logistic regression, especially at lower cutoffs, where deep neural networks perform especially poorly.

In order to understand what contributes to the false positives (e.g., 0.6 for precision@1 for post prediction) and false negatives, we further explored the reason behind misclassification of posts. More specifically, we filtered out the posts and friend-post pairs that were misclassified (false positives and false negatives) by our predictor by a significant margin. We then performed qualitative analysis on the participant-provided justification for their decisions about these posts' privacy settings to unpack possible rationales.

8.5 Analyzing Post Prediction Inaccuracy

Here, we qualitatively investigate the predictions missed by our classifier and provide a comprehensive analysis of these misclassified posts. We envision this analysis to be beneficial for future study designs by allowing researchers to gain insight into useful features to account for while building such automated learning tools. In addition, we also highlight the need for understanding personalized user contexts when designing such human-in-the-loop interfaces.

We perform this analysis on both the **post dataset** and the **friend-post dataset** predictions. We use the percentage of *limit sharing* choices in the training data as the cutoff k and aggregate all false positives and negatives across the 5 testing folds. For false negatives, we focus on suggestions ranked in the bottom 50% of the aggregated set as these are misclassified by a significant margin.

Post-based features
Details of content associated with a post (e.g., labeling images / video)
Classes of sensitive information within the post text or content
Similarity analysis of post content with the participant's present interests
Friend-based features
The interests, likes, and dislikes of the participant's friends
If particular friends are close family or otherwise related
Frequency of offline interaction between the participant and their friends

Table 5: Potential features to collect in future studies.

A fair number (42%) of misclassified posts were caused by the absence of accurate predictive features in our dataset. A significant number of these misclassified posts are linked to external content such as associated images, videos, or news articles. To ensure participants' privacy, and due to a lack of discussion in current related work about significant predictive features, we chose not to collect features specific to posts' external content. In other cases, participants' responses also suggest the presence of whole classes of sensitive content, e.g. *"I would like posts of my children to be as private as possible."* While we collect individual examples and reasons, sufficiently described classes of sensitive content would likely be a helpful supplement to our approach.

One additional source of inaccuracy was a lack of features specific to participants' friends. For 16% of misclassified friend-post pairs, participants mentioned the content of a given post being closely aligned with their particular friend's interests. For instance, one participant explained, *"I think she likes articles about animals."* There were also cases where participants mentioned that their friend would not like the content or it would be controversial. As our friend-based features do not account for the preferences of participants' friends and we did not attempt to collect this information for privacy and consent reasons, such instances are hard to predict.

Some misclassified posts were shared with close friends or family members with whom users wanted to continue sharing the posts. While Facebook allows participants to list family members on their profile, we did not collect this information. In other friend-post pairs, the level of interaction was not always representative of the closeness of their relationship and led to an inaccurate prediction. For instance, one participant said about a specific friend-post pair, *"He's a long distance boyfriend that I grew up with so I don't really care too much if he sees it or doesn't."* As the dynamics of Facebook and its users change, online interaction levels will not always be sufficient to determine complex social connections. Having access to additional complementary features (e.g., family relationships) can enable the development of more accurate classifiers.

In summary, elaborating on our findings from this investigative analysis on mispredictions, Table 5 presents a list of useful features that, if collected, could enable more accurate models for predicting privacy-setting misalignment in the future.

Our analysis also revealed the strong presence of personalized context, which limits the extent to which fully automated classifiers can predict an individual's preferences. For example, when explaining a change to the privacy setting of a post, a participant wrote, *"I no longer participate in these activities and don't find them appropriate any longer."* Inferring a connection between participation in an activity, its appropriateness, and a desired sharing setting may in fact be possible, but such nuanced and subjective connections are unlikely to be currently achievable. In other misclassified instances, participants' explanations emphasized the audience of a post. For example, one participant wrote, *"It was set to friends and that's the only people who I'd want to have my phone number."* Without access to preferences regarding explicitly curated sharing lists, developing an accurate understanding of friends' closeness in light of their limited social media interaction is non-trivial.

While the goal of any automated inference system is to minimize or eliminate inaccuracies, a domain as subjective and contextual as personal information sharing is bound to have occasional mistakes. When initially designing such a system, a human-centered investigation of the mental models and preferences regarding these decisions can provide valuable insights regarding what additional features to collect, as well as which inference rules may not accurately generalize across different individuals.

9 DISCUSSION AND CONCLUSIONS

For users, access control is typically a "set it, and forget it" endeavor. Even if the privacy setting a user has chosen for a social media post was accurate at the time it was set, it may be inappropriate moving forward. This mismatch can result from changes in the user's life and relationships, in addition to changes in the affordances and usage of the sharing platform itself. In our user study, we asked 78 Facebook users to evaluate five of their previous Facebook posts. For one-quarter of these posts, participants reported that they preferred to move forward with a privacy setting different from the one currently set. Participants wanted to reduce posts' audience sizes roughly as often as they want to increase them.

While we had initially hypothesized that one could predict which privacy settings ought to change based on how frequently participants interacted with particular friends or when they became Facebook friends, these characteristics had no predictive power for the task at hand. Participants desired to maintain sharing with low-interaction (but high-importance) classes of friends like family members. This insight is in line with previous work on invisible audiences [9, 48] and further highlights the importance of low-interaction friend connections on social networks.

In contrast, we showed promising results when building predictive models for users who wish to limit the privacy of past posts. Our results show that predicting the desired privacy settings of friend-post pairs is a particularly viable approach. We find that it is possible to automatically generate a ranked list of friend-post pairs for which the highest ranked pairs are likely to be cases for which the user wishes to retrospectively limit sharing for the post. Compared to baseline methods that consider the level of publicly visible interaction on Facebook, our predictive models perform more than three times better when identifying the friend-post pairs where the user would want to limit the audience. Additionally, when considering the most useful features in our predictive models, we found that focusing only on features that can be collected automatically (rather than requiring explicit user interaction) minimally impacts predictive performance. Thus, the initial identification of such friend-post pairs can proceed without burdening users.

Potential deployment: Privacy decisions are often nuanced and highly contextual. As our results on low-interaction, yet high-importance, Facebook friends illustrate, the data necessary to fully

contextualize a privacy decision may not even be available in the system in the first place. Furthermore, while our predictive models are successful at ranking friend-post pairs such that the highest ranked pairs are likely to require privacy reevaluation, the current versions of these models have insufficient accuracy for automatically determining privacy settings for all posts.

As a result, we imagine that our predictive models would be most successfully deployed as part of a human-in-the-loop interface. For example, similar to Facebook's "friends you may know" suggestion box, we imagine our classifier's highest-ranked suggestions being presented to the user as "posts whose privacy settings you may wish to revisit." Users could actively engage with these suggestions, evaluating them in terms of their unique knowledge outside the system (e.g., about their intended self-presentation and real-world relationships with the recipients). Because of this human-in-the-loop process, near-perfect prediction accuracy is not necessary. False positives generated by the classifier will be evaluated by the user, who will likely choose to keep the current privacy setting. While a high rate of false positives might discourage attention and engagement, our classifier results suggest that most of the highly ranked friend-post pairs are likely to be true positives. As a result of this human-in-the-loop aspect, the posts that are hidden based on the user's affirmative decisions are those they intend to hide.

When dealing with modern volumes of friend-post pairs for which to maintain proper privacy settings, our work demonstrates a promising approach to partially automating this process. This approach promises to focus the user's attention toward privacy settings that need to be revisited far better than requiring users to manually sift through past posts. Future work, however, is essential for further specifying and designing potential human-in-the-loop interfaces, as well as evaluating them in practice.

Low-interaction friends can be important: Our results highlight participants' desire to keep sharing with low-interaction, but high-importance, friends, such as family members. Any interaction-based cutoff for removing or reevaluating sharing decisions would incorrectly remove these connections. This insight is in line with previous work on invisible audiences [9, 48].

Additional external data can better contextualize posts: In the case of inaccuracies, the data needed to correctly classify posts was often not available through Facebook. Future research in this area can mine external (e.g., the content to which URLs point) and non-textual data (e.g., images, videos). At a high level, participant responses suggested that individuals intend to broadly share content of general interest (e.g., news and humor) while restricting the audience of personal content. When participants were asked why they wanted to change a given post's audience, they were far more likely to cite reasons related to the content of the post (e.g., "It's irrelevant because it's an old sports post about a game") rather than friendship dynamics or life events. Our qualitative coding of participants' self-reported reasoning leads us to believe that post content is an important determinant for whether a post's privacy setting should be changed. We combine this insight with two key reasons for our prediction inaccuracy — the presence of external content and limited text content — to suggest that future work analyzing post content more deeply is likely to better predict changes. Our qualitative results also indicated that privacy decisions were sometimes rooted in participants' anticipation of their friends' interests, contributing to prediction inaccuracies.

To protect participants' privacy, we restricted our analysis to data on our participants' Facebook accounts. Future work could include external data with proper consent, which is likely to further aid in identifying past posts in need of retrospective privacy management. Therefore, future work should focus on using additional data mined from connected URLs, as well as further analyzing images and videos. Our deep approach to investigating post privacy decisions provided useful insights that refined our intuition about how to operationalize retrospective tools. It is a natural precursor to a broader, quantitative approach to this task.

Limitations and future work: As with most studies conducted on real user data, our study has limitations. Because we wanted to probe deeply into several posts for individual participants, our overall sample size is lower than one might want for quantitative analysis. Furthermore, a likely nontrivial bias is introduced by the necessity of allowing our tools to investigate the full contents of the participant's Facebook account. This will likely dissuade privacy-sensitive users from participating in this or any other study of the same phenomenon when it requires informed consent.

Because we wanted to probe deeply into several posts for individual participants, our experimental approach is not well-suited for large-scale analysis. While not conclusive, our promising prediction results are hopefully a lower bound that will only improve with access to more training data. Leveraging qualitative insights, Table 5 highlighted additional features to collect in future studies. We envision this additional data will improve prediction accuracy.

While we found no simple mismatch between user preferences and current privacy settings that could be corrected in a fully automated way, we were able to make significant headway toward this high-level goal. By building a model founded on both qualitative and quantitative insights, we took a first step toward developing human-in-the-loop retrospective privacy-protection systems.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. CNS-1801663 and CNS-1351058. We thank the reviewers and our shepherd, Weili Han, for their feedback. We also thank Dimitri Vasilkov, William Wang, and Xuefeng Liu.

REFERENCES

[1] Alessandro Acquisti and Ralph Gross. 2006. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In *Proc. PETS*.
[2] Hazim Almuhimedi, Shomir Wilson, Bin Liu, Norman Sadeh, and Alessandro Acquisti. 2013. Tweets Are Forever: A Large-Scale Quantitative Analysis of Deleted Tweets. In *Proc. CSCW*.
[3] Oshrat Ayalon and Eran Toch. 2013. Managing Longitudinal Privacy in Online Social Networks. In *Proc. SOUPS*.
[4] Oshrat Ayalon and Eran Toch. 2017. Not Even Past: Information Aging and Temporal Privacy in Online Social Networks. *Human Compuer Interaction* 32, 2 (2017), 73–102.
[5] Lars Backstrom, Eytan Bakshy, Jon Kleinberg, Thomas Lento, and Itamar Rosenn. 2011. Center of Attention: How Facebook Users Allocate Attention across Friends. In *Proc. ICWSM*.
[6] Lujo Bauer, Lorrie Faith Cranor, Saranga Komanduri, Michelle L. Mazurek, Michael K. Reiter, Manya Sleeper, and Blase Ur. 2013. The Post Anachronism: The Temporal Dimension of Facebook Privacy. In *Proc. WPES*.
[7] Lujo Bauer, Lorrie Faith Cranor, Robert W. Reeder, Michael K. Reiter, and Kami Vaniea. 2009. Real Life Challenges in Access-Control Management. In *Proc. CHI*.

[8] Matthias Beckerle and Leonardo A Martucci. 2013. Formal Definitions for Usable Access Control Rule Sets From Goals to Metrics. In *Proc. SOUPS*.

[9] Michael S. Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. 2013. Quantifying the Invisible Audience in Social Networks. In *Proc. CHI*.

[10] Smriti Bhagat, Graham Cormode, Balachander Krishnamurthy, and Divesh Srivastava. 2010. Privacy in Dynamic Social Networks. In *Proc. WWW*.

[11] Will Brackenbury, Rui Liu, Mainack Mondal, Aaron Elmore, Blase Ur, Kyle Chard, and Michael J. Franklin. 2019. Draining the Data Swamp: A Similarity-based Approach. In *Proc. HILDA*.

[12] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. 2013. Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science* 4, 3 (2013), 340–347.

[13] Petter Bae Brandtzaeg and Marika Lüders. 2018. Time Collapse in Social Media: Extending the Context Collapse. *Social Media + Society* 4, 1 (2018).

[14] Xiang Cao and Lee Iverson. 2006. Intentional Access Management: Making Access Control Usable for End-Users. In *Proc. SOUPS*.

[15] Bernhard Debatin, Jennette P. Lovejoy, Ann-Kathrin Horn, and Brittany N. Hughes. 2009. Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Journal of Computer-Mediated Communication* 15, 1 (2009), 83–108.

[16] Serge Egelman, Andrew Oates, and Shriram Krishnamurthi. 2011. Oops, I Did it Again: Mitigating Repeated Access Control Errors on Facebook. In *Proc. CHI*.

[17] Facebook. 2018. App-Scoped IDs. https://developers.facebook.com/docs/apps/faq#app-scoped-ids. (Last accessed in August 2019).

[18] Facebook. 2018. How do I choose who can see previous posts on my timeline? https://www.facebook.com/help/236898969688346. (Last accessed in August 2019).

[19] Facebook. 2018. What audiences can I choose from when I share? https://www.facebook.com/help/211513702214269. (Last accessed in August 2019).

[20] Facebook. 2018. What is public information? https://www.facebook.com/help/203805466323736. (Last accessed in August 2019).

[21] Facebook. 2018. What's Privacy Checkup and how can I find it? https://www.facebook.com/help/443357099140264/. (Last accessed in August 2019).

[22] Facebook. 2018. When I post something, how do I choose who can see it? https://www.facebook.com/help/120939471321735. (Last accessed in August 2019).

[23] Lujun Fang and Kristen LeFevre. 2010. Privacy Wizards for Social Networking Sites. In *Proc. WWW*.

[24] Casey Fiesler, Michaelanne Dye, Jessica L. Feuston, Chaya Hiruncharoenvate, Clayton J. Hutto, Shannon Morrison, Parisa Khanipour Roshan, Umashanthi Pavalanathan, Amy S. Bruckman, Munmun De Choudhury, and Eric Gilbert. 2017. What (or Who) Is Public?: Privacy Settings and Social Media Content Sharing. In *Proc. CSCW*.

[25] Ronald A. Fischer. 1922. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society* 85, 1 (1922), 87—94.

[26] David Garcia. 2017. Leaking Privacy and Shadow Profiles in Online Social Networks. *Science Advances* 3, 8 (2017).

[27] Kambiz Ghazinour, Stan Matwin, and Marina Sokolova. 2013. Monitoring and Recommending Privacy Settings in Social Networks. In *Proc. EDBT*.

[28] Eric Gilbert and Karrie Karahalios. 2009. Predicting Tie Strength with Social Media. In *Proc. CHI*.

[29] Neil Zhenqiang Gong and Bin Liu. 2016. You Are Who You Know and How You Behave: Attribute Inference Attacks via Users' Social Friends and Behaviors. In *Proc. USENIX Security*.

[30] Google. 2019. Google Cloud Natural Language. https://cloud.google.com/natural-language/. (Last accessed in August 2019).

[31] Whitson Gordon. 2011. How to Use Facebook's New Timeline Feature (and Hide Your Embarrassing Old Posts). https://lifehacker.com/how-to-use-facebooks-new-timeline-feature-and-hide-you-5868411. (Last accessed in August 2019).

[32] Instagram. 2019. Stories. https://help.instagram.com/1660923094227526. (Last accessed in August 2019).

[33] Maritza Johnson, Serge Egelman, and Steven M. Bellovin. 2012. Facebook and Privacy: It's Complicated. In *Proc. SOUPS*.

[34] Jonathan Gheller. 2015. Introducing On This Day: A New Way to Look Back at Photos and Memories on Facebook. https://newsroom.fb.com/news/2015/03/introducing-on-this-day-a-new-way-to-look-back-at-photos-and-memories-on-facebook/. (Last accessed in August 2019).

[35] Patrick Gage Kelley, Robin Brewer, Yael Mayer, Lorrie Faith Cranor, and Norman Sadeh. 2011. An Investigation into Facebook Friend Grouping. In *Proc. INTERACT*.

[36] Patrick Gage Kelley, Paul Hankes Drielsma, Norman Sadeh, and Lorrie Faith Cranor. 2008. User-controllable Learning of Security and Privacy Policies. In *Proc. AISec*.

[37] Mohammad Taha Khan, Maria Hyun, Chris Kanich, and Blase Ur. 2018. Forgotten But Not Gone: Identifying the Need for Longitudinal Data Management in Cloud Storage. In *Proc. CHI*.

[38] Peter Klemperer, Yuan Liang, Michelle Mazurek, Manya Sleeper, Blase Ur, Lujo Bauer, Lorrie Faith Cranor, Nitin Gupta, and Michael Reiter. 2012. Tag, You Can See It!: Using Tags for Access Control in Photo Sharing. In *Proc. CHI*.

[39] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private Traits and Attributes are Predictable From Digital Records of Human Behavior. *PNAS* 110, 15 (2013), 5802–5805.

[40] Balachander Krishnamurthy and Craig E. Wills. 2008. Characterizing Privacy in Online Social Networks. In *Proc. WOSN*.

[41] Devdatta Kulkarni and Anand Tripathi. 2008. Context-Aware Role-based Access Control in Pervasive Computing Systems. In *Proc. SACMAT*.

[42] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174.

[43] Jack Lindamood, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. 2009. Inferring Private Information Using Social Network Data. In *Proc. WWW*.

[44] Heather Richter Lipford, Andrew Besmer, and Jason Watson. 2008. Understanding Privacy Settings in Facebook with an Audience View. In *Proc. UPSEC*.

[45] Eden Litt and Eszter Hargittai. 2016. The Imagined Audience on Social Network Sites. *Social Media + Society* (2016).

[46] Kun Liu and Evimaria Terzi. 2010. A Framework for Computing the Privacy Scores of Users in Online Social Networks. *TKDD* 5, 1 (2010), 6.

[47] Yabing Liu, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove. 2011. Analyzing Facebook Privacy Settings: User Expectations vs. Reality. In *Proc. IMC*.

[48] Mark Lochrie and Paul Coulton. 2012. Sharing the Viewing Experience Through Second Screens. In *Proc. EuroITV*.

[49] Michelle Madejski, Maritza Johnson, and Steven M. Bellovin. 2012. A Study of Privacy Settings Errors in an Online Social Network. In *Proc. PerCom*.

[50] Larry Magid. 2014. Facebook Changes New User Default Privacy Setting To Friends Only – Adds Privacy Checkup. Forbes https://www.forbes.com/sites/larrymagid/2014/05/22/facebook-changes-default-privacy-setting-for-new-users/. (Last accessed in August 2019).

[51] Michelle L. Mazurek, Peter F. Klemperer, Richard Shay, Hassan Takabi, Lujo Bauer, and Lorrie Faith Cranor. 2011. Exploring Reactive Access Control. In *Proc. CHI*.

[52] Michelle L. Mazurek, Yuan Liang, William Melicher, Manya Sleeper, Lujo Bauer, Gregory R. Ganger, Nitin Gupta, and Michael K. Reiter. 2014. Toward Strong, Usable Access Control for Shared Distributed Data. In *Proc. FAST*.

[53] Alessandra Mazzia, Kristen LeFevre, and Eytan Adar. 2012. The PViz Comprehension Tool for Social Network Privacy Settings. In *Proc. SOUPS*.

[54] Matt Mckeon. 2010. The Evolution of Privacy on Facebook. http://mattmckeon.com/facebook-privacy/. (Last accessed in August 2019).

[55] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proc. NIPS*.

[56] Brett Molina. May 9, 2017. Your snaps can now last to 'infinity' with Snapchat's latest update. *USA Today*.

[57] Mainack Mondal, Yabing Liu, Bimal Viswanath, Krishna P. Gummadi, and Alan Mislove. 2014. Understanding and Specifying Social Access Control Lists. In *Proc. SOUPS*.

[58] Mainack Mondal, Johnnatan Messias, Saptarshi Ghosh, Krishna P. Gummadi, and Aniket Kate. 2016. Forgetting in Social Media: Understanding and Controlling Longitudinal Exposure of Socially Shared Data. In *Proc. SOUPS*.

[59] Mainack Mondal, Johnnatan Messias, Saptarshi Ghosh, Krishna P. Gummadi, and Aniket Kate. 2017. Longitudinal Privacy Management in Social Media: The Need for Better Controls. *IEEE Internet Computing* 21, 3 (2017), 48–55.

[60] Karl Pearson. 1900. On the Criterion that a Given System of Deviations From the Probable in the Case of a Correlated System of Variables is Such That it Can Be Reasonably Supposed to Have Arisen From Random Sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50, 302 (1900), 157–175.

[61] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake VanderPlas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[62] Robert W. Reeder, Lujo Bauer, Lorrie Faith Cranor, Michael K. Reiter, Kelli Bacon, Keisha How, and Heather Strong. 2008. Expandable Grids for Visualizing and Authoring Computer Security Policies. In *Proc. CHI*.

[63] Thomson Reuters. 2018. Reuters Poll Data. https://fingfx.thomsonreuters.com/gfx/rngs/FACEBOOK-PRIVACY-POLL/010062SJ4QF/2018%20Reuters%20Tracking%20-%20Social%20Media%20Usage%205%203%202018.pdf. (Last accessed in August 2019).

[64] Ravi S. Sandhu, Edward J. Coyne, Hal L. Feinstein, and Charles E. Youman. 1996. Role-based Access Control Models. *Computer* 29, 2 (1996), 38–47.

[65] Selenium. 2018. Selenium browser automation. https://www.seleniumhq.org/. (Last accessed in August 2019).

[66] Evan Selinger and Woodrow Hartzog. 2016. Facebook's Emotional Contagion Study and the Ethical Problem of Co-opted Identity in Mediated Environments Where Users Lack Control. *Research Ethics* 12, 1 (2016), 35–43.
[67] Richard T. Simon and Mary Ellen Zurko. 1997. Separation of Duty in Role-based Environments. In *Proc. CSF*.
[68] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. 2013. The Post That Wasn't: Exploring Self-censorship on Facebook. In *Proc. CSCW*.
[69] Manya Sleeper, Justin Cranshaw, Patrick Gage Kelley, Blase Ur, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. 2013. "I Read My Twitter the Next Morning and Was Astonished": A Conversational Perspective on Twitter Regrets. In *Proc. CHI*.
[70] Diana K. Smetters and Nathan Good. 2009. How Users Use Access Control. In *Proc. SOUPS*.
[71] Statista. 2018. Distribution of Facebook users in the United States as of January 2018, by gender. https://web.archive.org/web/20181116070219/https://www.statista.com/statistics/266879/facebook-users-in-the-us-by-gender/. (Last accessed in August 2019).
[72] Statista. 2018. Number of Facebook users by age in the U.S. as of January 2018 (in millions). https://www.statista.com/statistics/398136/us-facebook-user-age-groups/. (Last accessed in August 2019).
[73] Fred Stutzman, Ralph Gross, and Alessandro Acquisti. 2013. Silent Listeners: The Evolution of Privacy and Disclosure on Facebook. *Journal of Privacy and Confidentiality* 4, 2 (2013), 7–41.
[74] Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29, 1 (2010), 24–54.
[75] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A Field Trial of Privacy Nudges for Facebook. In *Proc. CHI*.
[76] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. "I Regretted the Minute I Pressed Share": A Qualitative Study of Regrets on Facebook. In *Proc. SOUPS*.
[77] WeChat Help Center. 2019. How do I set permissions for Moments? https://help.wechat.com/cgi-bin/micromsg-bin/oshelpcenter?opcode=2&plat=android&lang=en&id=120813euEJVf141023RBfMjm. (Last accessed in August 2019).
[78] Pamela Wisniewski, Bart P. Knijnenburg, and Heather Richter Lipford. 2017. Making Privacy Personal: Profiling Social Network Users to Inform Privacy Education and Nudging. *International Journal of Human-Computer Studies* 98 (2017), 95–108.
[79] XGBoost Developers. 2016. XGBoost. https://xgboost.readthedocs.io/en/latest/.
[80] Haochen Zhang, Min-Yen Kan, Yiqun Liu, and Shaoping Ma. 2014. Online Social Network Profile Linkage. In *Proc. AIRS*.
[81] Elena Zheleva and Lise Getoor. 2007. Preserving the Privacy of Sensitive Relationships in Graph Data. In *Proc. PinKDD*.
[82] Elena Zheleva and Lise Getoor. 2009. To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. In *Proc. WWW*.
[83] Elena Zheleva, Evimaria Terzi, and Lise Getoor. 2012. Privacy in Social Networks. *Synthesis Lectures on Data Mining and Knowledge Discovery* 3, 1 (2012), 1–85.
[84] Mary Ellen Zurko, Rich Simon, and Tom Sanfilippo. 1999. A User-Centered, Modular Authorization Service Built on an RBAC Foundation. In *Proc. IEEE S&P*.

APPENDIX

A SURVEY 1 QUESTIONS

Longitudinal Privacy Behaviors

First, we would like to ask you about how you use this account to connect with family. I use this Facebook account primarily for the following purposes. Select all that apply. ◯ Sending direct messages to family (e.g., via Facebook Messenger) ◯ Looking through the newsfeed to stay up to date with family ◯ Liking, sharing, or commenting on things my family posted ◯ Sharing pictures with family ◯ Writing text posts (e.g., status updates) for family ◯ Sharing content posted by others (e.g., new articles, links) for family ◯ Other: ___ ◯ None of the above

Next, we would like to ask you about how you use this account to connect with close friends. I use this Facebook account primarily for the following purposes. Select all that apply. ◯ Sending direct messages to close friends (e.g., via Facebook Messenger) ◯ Looking through the newsfeed to stay up to date with close friends ◯ Liking, sharing, or commenting on things my close friends posted ◯ Sharing pictures with close friends ◯ Writing text posts (e.g., status updates) for close friends ◯ Sharing content posted by others (e.g., new articles, links) for close friends ◯ Other: ___ ◯ None of the above

Next, we would like to ask you about how you use this account to connect with professional contacts. I use this Facebook account primarily for the following purposes. Select all that apply. ◯ Sending direct messages to professional contacts (e.g., via Facebook Messenger) ◯ Looking through the newsfeed to stay up to date with professional contacts ◯ Liking, sharing, or commenting on things my professional contacts posted ◯ Sharing pictures with professional contacts ◯ Writing text posts (e.g., status updates) for professional contacts ◯ Sharing content posted by others (e.g., new articles, links) for professional contacts ◯ Other: ___ ◯ None of the above

Finally, we would like to ask you about how you use this account to connect with acquaintances (e.g. people you meet briefly at an event). I use this Facebook account primarily for the following purposes. Select all that apply. ◯ Sending direct messages to acquaintances (e.g., via Facebook Messenger) ◯ Looking through the newsfeed to stay up to date with acquaintances ◯ Liking, sharing, or commenting on things acquaintances posted ◯ Sharing pictures with acquaintances ◯ Writing text posts (e.g., status updates) for acquaintances ◯ Sharing content posted by others (e.g., new articles, links) for acquaintances ◯ Other: ___ ◯ None of the above

Do you have any other Facebook accounts? ◯ Yes, for the purposes of ___ ◯ No

Do you use any of the following social media platforms? For each one, rank from 1 (I do not use this platform) to 4 (I use this platform more frequently than I use Facebook)

- Twitter
- Instagram
- Snapchat
- Reddit
- YouTube
- Tumblr
- LinkedIn
- WhatsApp
- Facebook Messenger
- Skype

In which year do you think the amount of time you spend on Facebook peaked? ___

Compared to the year when my Facebook usage peaked, I currently use Facebook: ◯ about as frequently as during that year ◯ a little less frequently ◯ much less frequently

Are you friends with any members of your immediate family on Facebook? ◯ Yes ◯ No ◯ I'm not sure

Are you friends with any members of your extended family on Facebook? ◯ Yes ◯ No ◯ I'm not sure

Are you friends with any of your work colleagues on Facebook? ◯ Yes ◯ No ◯ I'm not sure

Are you friends with any people you went to school with (at any level, from grade school through graduate school) on Facebook? ◯ Yes ◯ No ◯ I'm not sure

Are you friends with acquaintances (e.g. people you meet briefly at an event) on Facebook? ◯ Yes ◯ No ◯ I'm not sure

Are you friends with anyone you have not met in person on Facebook? ◯ Yes ◯ No ◯ I'm not sure

Do you use Facebook for work-related purposes? ◯ Yes, I use Facebook in order to ___ ◯ No

I consider this Facebook account essential to my personal social life. ◯ Strongly agree ◯ Somewhat agree ◯ Neither agree nor disagree ◯ Somewhat disagree ◯ Strongly disagree

I consider this Facebook account essential to my professional life. ◯ Strongly agree ◯ Somewhat agree ◯ Neither agree nor disagree ◯ Somewhat disagree ◯ Strongly disagree

In the past, have you ever unfriended any of your Facebook friends? ◯ Yes ◯ No

If yes, what are some reasons you unfriended people in the past? You can give more than 1 reason. ___

If no, have you ever considered unfriending any of your Facebook friends? If the answer is yes, why did you not unfriend them in the end? ___

Have you ever gone back and changed the audience that can view a post after you initially posted it? ◯ Yes ◯ No

If yes, what are some reasons why you change the audience of (an) old post(s)? ___

If no, have you ever considered changing the audience of a past post? If the answer is yes, why did you not change the audience in the end? ___

Please think back to one year after you created your Facebook account. At that time, what did you often post about on Facebook? ___

Please describe how you used Facebook in general one year after you created your Facebook account. That is, what did you use it for, and what was your approach to Facebook? ___

This question asks about the midpoint of your Facebook account. If you created your account in 2010, for example, the midpoint between 2010 in 2018 would be 2014. As of the midpoint of your Facebook account, what did you often post about on Facebook? ___

This question asks about the midpoint of your Facebook account. If you created your account in 2010, for example, the midpoint between 2010 in 2018 would be 2014. Please describe how you used Facebook in general as of the midpoint of your Facebook account. ___

At the present time, what do you often post about on Facebook? ___

At the present time, please describe how do you use Facebook in general. ___

To your knowledge, have any significant events or changes in your personal life changed how you decide what to share on Facebook? ◯ Yes ◯ No

If yes, please briefly describe one significant event or change in your personal life which impacted how you decide what to share on Facebook. ___

If yes, approximately when did this event or change in your personal life happen? (How many months or years ago?) ___

If yes, how did this event or change in your personal life impact how you decide what to share on Facebook? Why? ___

To your knowledge, have any significant events or changes in your professional life changed how you decide what to share on Facebook? ◯ Yes ◯ No

If yes, please briefly describe one significant event or change in your professional life which impacted how you decide what to share on Facebook. ___

If yes, approximately when did this event or change in your professional life happen? (How many months or years ago?) ___

If yes, how did this event or change in your professional life impact how you decide what to share on Facebook? Why? ___

To your knowledge, have any news stories or events concerning either Facebook or the world more broadly impacted how you decide what to share on Facebook?

If yes, please briefly describe one news story or event that impacted how you decide what to share on Facebook. ___

If yes, in what way did this news story or event impact how you decide what to share on Facebook? Why? ___

Did the way you decide what to share on Facebook change at all since you started using the platform? ◯ Yes, due to ___ ◯ No, because ___

If yes, how did the way you decide what to share on Facebook change since you started using the platform? ___

Today, I frequently adopt strategies to protect my privacy on Facebook. ◯ Strongly agree ◯ Somewhat agree ◯ Neither agree nor disagree ◯ Somewhat disagree ◯ Strongly disagree

In the first year I started using Facebook, I frequently adopted strategies to protect my privacy on Facebook. ◯ Strongly agree ◯ Somewhat agree ◯ Neither agree nor disagree ◯ Somewhat disagree ◯ Strongly disagree

I would expect that ___ browsed my Facebook profile in the past to find an old post. ◯ none of my friends ◯ some of my friends ◯ most of my friends ◯ all of my friends

How would you feel about a Facebook friend browsing your profile to look at posts that are at least one year old? ___

Why would you expect they would do this? ___

What kind of posts that are at least one year old would you expect other people might look at? ___

How would you feel about a Facebook friend browsing your profile to look at posts that are at least three years old? ___

Why would you expect they would do this? ___

What kind of posts that are at least three years old would you expect other people might look at? ___

Have you ever browsed a friend's Facebook profile in order to look at posts that are at least one year old (at the time of your search)? ◯ Yes ◯ No

If yes, why? ___

If yes, from looking at posts of that age on your friends' Facebook accounts, did you encounter any posts that surprised you? ◯ Yes ◯ No ◯ I'm not sure

If yes, can you describe what one of those posts was about in a sentence? ___

If you have never browsed a friend's Facebook profile in order to look at posts that are at least one year old, what kind of posts that are at least one year old from your friends would you consider looking at? ___

Have you ever browsed a friend's Facebook profile in order to look at posts that are at least three years old (at the time of your search)? ◯ Yes ◯ No

If yes, why? ___

If yes, from looking at posts of that age on your friends' Facebook accounts, did you encounter any posts that surprised you? ◯ Yes ◯ No ◯ I'm not sure

If yes, can you describe what one of those posts was about in a sentence? ___

If you have never browsed a friend's Facebook profile in order to look at posts that are at least three years old, what kind of posts that are at least three years old from your friends would you consider looking at? ___

Do you ever look back at things you posted on Facebook in the past? ◯ Yes ◯ No

If yes, why? ___

If no, why not? ___

Do you ever look back at things your friends have posted on your Facebook timeline in the past? ◯ Yes ◯ No

If yes, why? ___

If no, why not? ___

External Stimuli - Privacy Demo
Have you ever seen Facebook's Privacy Checkup feature before? ◯ Yes ◯ No ◯ I'm not sure

Have you ever used this feature before? ◯ Yes ◯ No ◯ I'm not sure

If yes, why did you use this feature? ___

If yes, from what you recall, what did you change by using this feature? ___

If you've seen this feature but didn't use it, why didn't you use this feature? ___

If you didn't use this feature, what would you expect this feature to do? ___

If you're not sure if you've used this feature, what would you expect this feature to do? ___

Have you ever seen Facebook's "limit the audience for all past posts" feature? ◯ Yes ◯ No ◯ I'm not sure

Have you ever used Facebook's "limit the audience for all past posts" feature? ◯ Yes ◯ No ◯ I'm not sure

If yes, why did you use this feature? ___

If you've seen this feature but didn't use it, why didn't you use this feature? ___

If you're not sure if you've used this feature, what would you expect this feature to do? ___

Demographics
With what gender do you identify? ◯ Male ◯ Female ◯ Non-binary ◯ Other ___ ◯ Prefer not to answer

What is your age? ◯ 18-24 ◯ 25-34 ◯ 35-44 ◯ 45-54 ◯ 55-64 ◯ 65-74 ◯ 75 or older ◯ prefer not to answer

Please specify your ethnicity. (Choose all that apply) ◯ White ◯ Hispanic or Latino ◯ Black or African American ◯ Native American or American Indian ◯ Asian / Pacific Islander ◯ Other ___ ◯ Prefer not to answer

What is the highest level of degree or level of school you have completed? ◯ No high school no diploma ◯ High school diploma ◯ Some college credit no degree ◯ Associate's degree (AA / AS) ◯ Bachelor's degree (BA / BS) ◯ Master's degree (MA, MS, MEd, MBA) ◯ Doctorate, Professional, or Terminal Degree (MD, PhD, DDS, DVM, JD, EdD) ◯ Prefer not to answer

What is your employment status? ◯ Student ◯ Full-time employed ◯ Part-time employed ◯ Not employed ◯ Retired ◯ Prefer not to answer

Are you majoring in or do you have a degree or job in computer science, computer engineering, information technology, or a related field? ◯ Yes ◯ No ◯ Prefer not to answer

B SURVEY 2 QUESTIONS

Content Specific Privacy Settings Questions
We chose 5 posts from the participant's past, and asked them the following questions about each post:

Prior to this survey, have you ever changed the sharing setting of this post? ◯ Yes ◯ No ◯ I considered changing the sharing setting, but ended up not changing it ◯ I'm not sure ◯ Prefer not to answer

If yes, approximately in what year did you change or considered changing the sharing setting? ___

Moving forward, ideally what sharing setting would you want to have for this post? ◯ I would want to have the current privacy setting ◯ Public ◯ Friends ◯ Friends except ___ ◯ Friends of friends ◯ Specific friends ___ ◯ Only me ◯ Custom (specify friends and lists you would like to include and/or exclude) ___ ◯ Delete this post from Facebook ◯ Prefer not to answer

How important is it that the existing privacy setting of the post be replaced by the new privacy setting that you just specified? ◯ Extremely important ◯ Very important ◯ Moderately important ◯ Slightly important ___ ◯ Not at all important ◯ N/A (I didn't mean to indicate a change in sharing setting) ◯ Prefer not to answer

If you wanted to change the privacy setting, why did you want to do so? ___

If you wanted to keep the same privacy setting, why did you want to do so? ___

For each of the 5 posts, we then asked the participant about their privacy preferences for the post with respect to 6 of their Facebook friends.

Indicate whether today you would want to keep sharing this post with this friend, stop sharing it with this friend, or whether it doesn't matter to you. ◯ Definitely keep sharing ◯ Probably keep sharing ◯ Doesn't matter ◯ Probably stop sharing ◯ Prefer not to answer

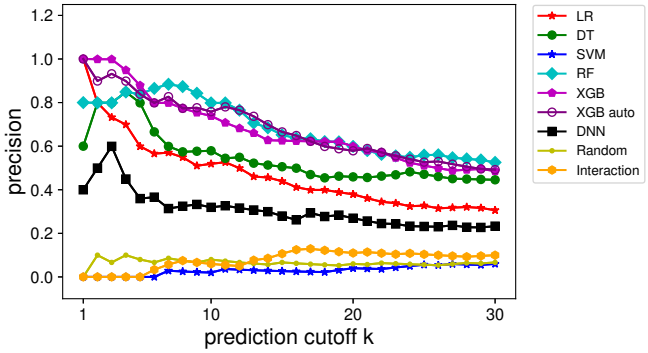
Please explain why. ___

If you chose definitely or probably stop sharing, would you want to friend to not be able to see ◯ this particular post only ◯ a number of my posts, including this post ◯ any of my posts ◯ Prefer not to answer
I consider the friend to be a close friend.◯ Strongly agree ◯ Somewhat agree ◯ Neither agree not disagree ◯ Somewhat disagree ◯ Strongly disagree ◯ Prefer not to answer

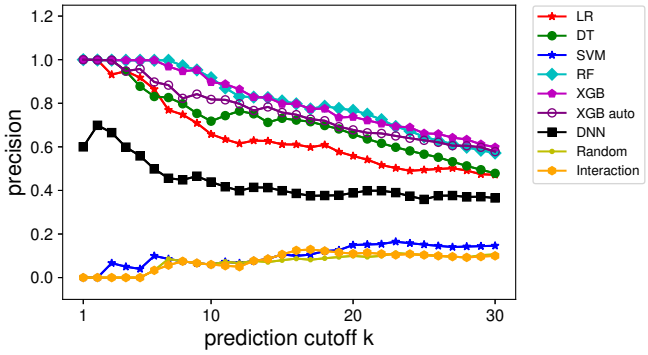
With the ideal privacy settings specified before, would this friend be able to see this post? ◯ Yes ◯ I'm not sure ◯ No ◯ Prefer not to answer
With the current privacy settings for this post, would this friend be able to see this post? ◯ Yes ◯ I'm not sure ◯ No ◯ Prefer not to answer

C ADDITIONAL FIGURES

This appendix presents a series of supplementary graphs and tables. First, we present four additional graphs for our predictive models, showing the relative performance of the different classifiers we tested on the same data. We then present two graphs analogous to those in the paper showing temporal patterns in the number of Facebook friends per participant and the privacy settings of posts. Different from the analogous graphs in the paper that included only the cohort whose accounts were at least a decade old, these variants show data from all participants. Finally, we include Table 6, which details the features we extracted from our survey or programmatically collected and then leveraged in our prediction task. These features are divided into four categories. In each category we consider multiple features.



(a) Average Precision@k on full test set



(b) Average Precision@k on test without label "doesn't matter"

Figure 10: Precision@k curves for the friend-post dataset, comparing the original dataset and removing the label *doesn't matter* from the testing folds. The preference for doesn't matter is most of the interference for the precision@k curves. We compare Logistic Regression (*LR*), Decision Trees (*DT*), Support Vector Machines (*SVM*), Random Forests (*RF*), XGBoost (*XGB*), Deep Neural Networks (*DNN*), random assignment (*Random*), and an interaction-based model (*Interaction*), as detailed in Section 8.1.

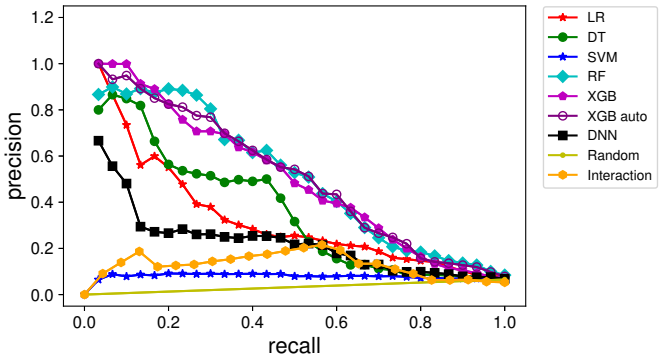


Figure 11: Precision vs. recall for the friend-post dataset.

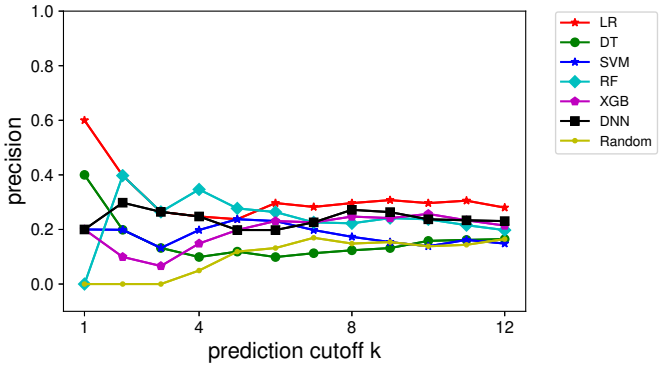


Figure 12: Average precision@k for post dataset.

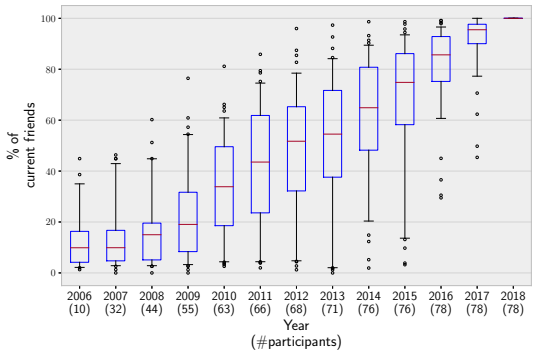


Figure 13: The percentage of participants' 2018 Facebook friends who were their Facebook friends in the past. The number in parentheses indicates how many participants had an account in that year.

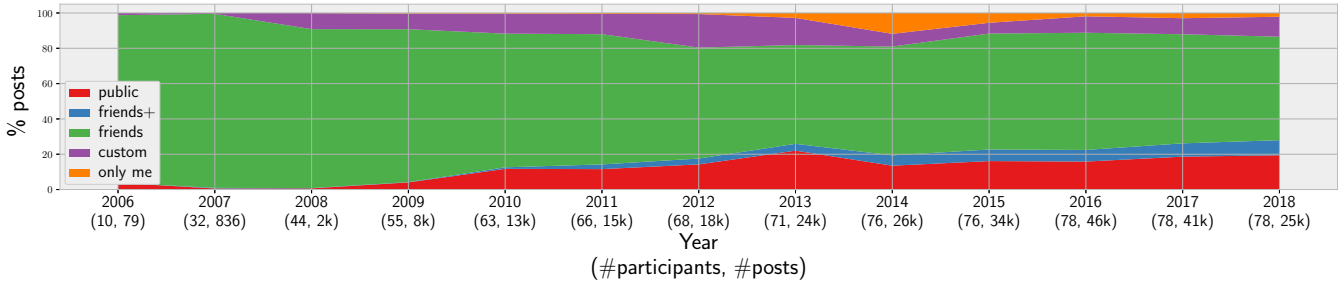


Figure 14: Staring from 2006, the yearly percentage of participants' posts shared with different privacy settings. Each participant's set of posts take up an equal amount of space on the y-axis. The majority of posts for all years are "friends only."

Category	Feature
Account (user) features	The age of the account (in years)
	Whether the participant had used Facebook's Privacy Check-Up
	Whether they had changed any post's privacy settings
	Whether they had ever unfriended a Facebook friend
	Whether personal life events impacted their sharing
	Whether professional life events impacted their sharing
	Whether news stories impacted their sharing
	The participant's age range
	Whether the participant had a CS or IT background
Post statistics based features	The age of the post (in years)
	# of likes, reactions, and comments (summed)
	The type of the post (e.g., text, photo)
	Whether the post contained a third-party link
	The post's current privacy setting (e.g., friends, public)
	Whether there is at least one comment on the post
	Whether there is at least one edited comment on the post
	Where another user is tagged in the post
Content based features	Whether the post text contains words from the LIWC categories (e.g., religious, swear, anger etc.) [74]. We obtained 63 categories for our dataset; each category corresponds to one feature. We used one-hot encoding to obtain binary feature values
	Whether the post text is classified into any of the Google content-classification categories (e.g., arts, politics, culture and entertainment) [30]. We obtained 21 categories for our dataset; each category correspond to one feature. We used one-hot encoding to obtain binary feature values
	Sentiment score of the post text computed by the Google Cloud Natural Language engine [30]
	Google News Word2Vec embedddings [55] of the post text
Audience based features (for specific friends)	# of days since first communication
	# of days since last communication
	# of days between first and last communication
	# of friends of the participant
	# of wall posts exchanged between the participant and the friend (we also used a normalized version as a separate feature)
	# of words exchanged via wall posts and comments (we also used a normalized version as a separate feature)
	# of intimate [74] wall words exchanged (we also used a normalized version as a separate feature)
	The number of likes and reactions that the participant gave on the friend's wall posts or the friend gave on the participant's wall posts (we also used a normalized version as a separate feature)

Table 6: A detailed enumeration of all features we used in our predictive model.