unFriendly: Multi-Party Privacy Risks in Social Networks

Kurt Thomas[1], Chris Grier[2], and David M. Nicol[1]

[1] University of Illinois at Urbana-Champaign {kathoma2,dmnicol}@illinois.edu
[2] University of California, Berkeley grier@cs.berkeley.edu

Abstract. As the popularity of social networks expands, the information users expose to the public has potentially dangerous implications for individual privacy. While social networks allow users to restrict access to their personal data, there is currently no mechanism to enforce privacy concerns over content uploaded by other users. As group photos and stories are shared by friends and family, personal privacy goes beyond the discretion of what a user uploads about himself and becomes an issue of what every network participant reveals. In this paper, we examine how the lack of joint privacy controls over content can inadvertently reveal sensitive information about a user including preferences, relationships, conversations, and photos. Specifically, we analyze Facebook to identify scenarios where conflicting privacy settings between friends will reveal information that at least one user intended remain private. By aggregating the information exposed in this manner, we demonstrate how a user's private attributes can be inferred from simply being listed as a friend or mentioned in a story. To mitigate this threat, we show how Facebook's privacy model can be adapted to enforce multi-party privacy. We present a proof of concept application built into Facebook that automatically ensures mutually acceptable privacy restrictions are enforced on group content.

1 Introduction

In the last decade the popularity of online social networks has exploded. Today, sites such as Facebook, MySpace, and Twitter combined reach over 500 million users daily [1–3]. As the popularity of social networks continues to grow, concerns surrounding sharing information online compound. Users regularly upload personal stories, photos, videos, and lists of friends revealing private details to the public. To protect user data, privacy controls have become a central feature of social networking sites [4, 5], but it remains up to users to adopt these features.

The sheer volume of information uploaded to social networks has triggered widespread concern over security and privacy [6, 7]. Personal data revealed on social networks has been used by employers for job screening [8] and by local law enforcement for monitoring and implicating students [9]. More sophisticated applications of social network data include tracking user behavior [10] and government funded monitoring [11]. Criminals have also capitalized on the trust users place in social networks, exploiting users with phishing attacks and malicious downloads [12, 13].

The diverse set of threats posed to users has resulted in a number of refinements to privacy controls [14]. However, one aspect of privacy remains largely unresolved: friends. As photos, stories, and data are shared across the network, conflicting privacy requirements between friends can result in information being unintentionally exposed to the public, eroding personal privacy. While social networks allow users to restrict access to their own data, there is currently no mechanism to enforce privacy concerns over data uploaded by other users. As social network content is made available to search engines [15] and mined for information [16], personal privacy goes beyond what one user uploads about himself; it becomes an issue of what every member on the network says and shares.

In this paper, we examine how the lack of *multi-party privacy* controls for shared content can undermine a user's privacy. We begin by analyzing situations in Facebook where asymmetric privacy requirements between two friends inadvertently weaken one user's privacy. This results in friends, tagged content, and conversations being unintentionally exposed to the public and crawlers. Using our examples as a foundation, we develop a formal definition of *privacy conflicts* to explore both the frequency and risk of information leaked by friends which cannot be prevented with existing privacy controls.

The presence of privacy conflicts between friends results in scattered references about a user appearing to the public, including being mentioned in a story, listed as a friend, or tagged in a photo. While a single conflict may pose a minimal risk to privacy, we show how the aggregate data revealed by conflicts can be analyzed to uncover sensitive information. We develop a classification system that uses publicly disclosed links between friends and the content of leaked conversations to build predictions about a user's gender, religious views, political leaning, and media interests. While predicting personal attributes based on friends has previously been examined [17–20], we present refinements to these techniques that utilize auxiliary information about mutual friends and the frequency and content of conversations to produce more accurate results. Our techniques highlight how various leaks of seemingly innocuous data can unintentionally expose meaningful private data, eroding personal privacy.

Using a data set of over 80,000 Facebook profiles, we analyze the frequency of asymmetric privacy requirements between friends, uncovering millions of instances where one user may potentially violate another user's privacy. We then process the aggregate information exposed by conflicts with our data analytic techniques, finding we are able to predict a user's personal attributes with up to 84% accuracy by simply using references and conversations exposed by friends.

To mitigate the threat of privacy conflicts, we show how the current Facebook privacy model can be adapted to enforce multi-party privacy. We present two proof of concept applications built into Facebook. One application simulates Facebook's popular wall functionality, while the other simulates a user's list of friends. The applications automatically determine a mutually acceptable privacy policy between groups of friends, only displaying information that all parties agree upon. Policy arbitration and enforcement is completely transparent to users, removing the risk of privacy conflicts without requiring user intervention.

2 Background and Motivation

Before describing the limitations of privacy in social networks, we present a brief overview of privacy controls currently available to users. While the prospect of friends and family weakening a user's privacy exists in all social networks, we restrict our analysis to Facebook given its status as the largest network with over 400 million users [1].

Facebook provides each user with a profile consisting of a page containing personal information, a list of the user's friends, and a wall where friends can post comments and leave messages, similar to a blog. A typical profile will contain information pertaining to the user's gender, political views, work history, and contact information. Additionally, users can upload stories, photos, and videos and *tag* other Facebook members that appear in the content. Each tag is an unambiguous reference that links to another user's profile, allowing a crawler to easily distinguish between Bob, Alice's friend and Bob, Carol's friend.

Privacy restrictions form a spectrum between public and private data. On the public end, users can allow every Facebook member to view their personal content. On the private end, users can restrict access to a specific set of trusted users. Facebook uses friendship to distinguish between trusted and untrusted parties. Users can allow *friends*, *friends of friends*, or *everyone* to access their profile data, depending on their personal requirements for privacy.

Despite the spectrum of available privacy settings, users have no control over information appearing outside their immediate profile page. When a user posts a comment to a friend's wall, he cannot restrict who sees the message. Similarly, if a user posts a photo and indicates the name of a friend in the photo, the friend cannot specify which users can view the photo. For both of these cases, Facebook currently lacks a mechanism to satisfy privacy constraints when more than one user is involved. This leads to *privacy conflicts*, where asymmetric privacy requirements result in one user's privacy being violated. Privacy conflicts publicly expose personal information, slowly eroding a user's privacy.

3 Multi-Party Privacy

To understand the risks posed by the lack of joint privacy controls in social networks, we construct a formalism for privacy conflicts that defines the situations where a user's privacy can be violated and the extent of information leaked. To develop this formalism, we begin by analyzing scenarios in Facebook where users can unintentionally violate one another's privacy. We then deconstruct these examples into a formalism that captures all potential privacy conflicts. This formalism plays an important role in Section 4 where we examine how information leaked by privacy conflicts can be analyzed to infer a user's personal attributes and in Section 6 where we show how Facebook can be adapted to enforce multi-party privacy.

3.1 Exploring Privacy Conflicts

Social networks are inherently designed for users to share content and make connections. When two users disagree on whom content should be exposed to,

we say a *privacy conflict* occurs. Multiple privacy conflicts can occur between a user and his friends, each revealing a potentially unique sensitive detail. We specifically analyze two scenarios in Facebook — friendship and wall posts — to understand the types of information exposed by conflicts.

Friendship: A central feature of social networks is the ability of users to disclose relationships with other members. Each relationship carries potentially sensitive information that either user may not wish revealed. While Facebook provides a mechanism to conceal a user's list of friends, the user can only control one direction of an inherently bidirectional relationship.

Consider a scenario where a user Alice adopts a policy that conceals all her friends from the public. On the other hand, Bob, one of Alice's friends, adopts a weaker policy that allows any user to view his friends. In this case, Alice's relationship with Bob can still be learned through Bob. We say that a privacy conflict occurs as Alice's privacy is violated by Bob's weaker privacy requirements.

Wall Posts and Tagging: Wall posts and status updates provide users with a built-in mechanism to communicate and share comments with other users. Each post consists of a sender, receiver, and the content to be displayed. Facebook currently allows only the receiver to specify a privacy policy. When Alice leaves a message on Bob's wall, she relinquishes all privacy control over her comments. Similarly, if Alice posts to her own wall, she has sole control over who can view the message, even if she references other users who wish to remain anonymous. By ignoring the privacy concerns of all but one user, information can be exposed that puts other friends at risk.

Consider an example where Alice makes a public comment on her own profile stating *"Skipping work with @Bob and hitting the bars at 9am"*. Bob is unambiguously identified by the message, but cannot specify that the message should not be broadcast to the public per his privacy policy. Alternatively, if Alice posts on Bob's profile about current relationship trouble, she cannot specify that the message should only be visible by her friends, not all of Facebook.

Additional Conflicts: Friendship and wall posts represent only two of numerous situations where Facebook and other social networks lack multi-party privacy. Group membership, fan pages, event attendance, photo tagging, and video tagging are additional situations where multiple parties can be referenced by data, but cannot control its exposure. Each exposure leaks sensitive information about a user even if the strictest privacy controls available are adopted.

3.2 Formalizing Privacy Conflicts

We now formalize multi-party privacy, creating a language to understand how existing privacy controls can still lead to undesired exposures. Consider a single social network user u in the set of all possible users U. We denote the pages owned by u such as the user's wall or friend list as the set G_u. For each page $g \in G_u$, the user u can specify a privacy policy $P_u(g)$ indicating set of users including u who can view the page. For instance, Alice can create a policy stating *everyone* can view her wall page. Here, u is Alice, g is the wall page, and $P_u(g)$ is the set

of all of users $u \in U$. We call the policy $P_u(g)$ the *owner policy*, as Alice controls access to the data and can remove it at any time.

Each page $g \in G_u$ contains a grouping of information I which may uniquely reference one or more users represented by the set $S(I)$. Here, Alice tagging Bob and Carol in a wall post i can be represented by $S(i) = \{Bob, Carol\}$. In this case, I is the set of all wall posts on the wall page g.

While the owner u of a page specifies the access restriction $P_u(g)$, each user referenced in the page will have a separate, potentially distinct privacy policy. For instance, while Alice may allow all users to view her wall page, Bob may desire all references of him be visible only to his direct friends. To capture this variation, we say that for each user $w \in S(I)$ there exists an *exposure policy* $V_w(g, I)$ that specifies a set of users permitted by w to view references in I about w on page g. This allows both an owner and exposed user to specify a policy for how data should be accessed, even if their policies are different. The lack of exposure policies in existing social networks is what allows information to be disseminated against a user's will.

We state that a *privacy conflict* occurs between the owner u of a page g and the users $S(I)$ referenced by the page if:

$$\exists i \in I : P_u(g) \nsubseteq \bigcap_{w \in S(i)} V_w(g, i) \tag{1}$$

That is to say, if an owner policy allows any users other than those accepted by *all* exposure policies to view a piece of information $i \in I$, there is at least one exposure policy being violated on page g. Returning to our example, Alice's owner policy $P_u(g) = U$ allows all users to view her wall page. This is in direct conflict with Bob's exposure policy $V_w(g, I) \subset U$ which requires his posts to be accessible only to his friends, not all users. Conversely, if Carol adopts an exposure policy $V_w(g, I) = U$, then Alice and Carol are in agreement on the set of users who can view the the information I on page g and no privacy conflict exists.

An important consequence of Equation 1 is that as the number of users referenced by a piece of information increases, in the absence of mutual friends, the intersection of all exposure policies tends to the empty set. This implies that for photos or wall posts referencing multiple users, it is likely that at least one user is being exposed against their will to undesired parties.

Currently, Facebook and other social networks lack a mechanism to specify an exposure policy. Instead, we can derive these policies based on the owner policy of each user. If Alice allows everyone to view her wall posts, her exposure policy is the same; all references to her in other wall posts should be visible to everyone. By using the formalism of owner policies and exposure policies, we can systematically examine Facebook to identify privacy conflicts and show how these violations can expose sensitive information.

3.3 Formalizing Exposed Data

Using our formalism of privacy conflicts, we can identify the set of all information pertaining to a particular user w that violates w's exposure policy. We denote

this set $E(w)$ which contains all Facebook pages including friendships, wall posts, and tags that leak information about w. We define $E(w)$ as:

$$E(w) = \{\forall(u \in U, g \in G, i \in I) : P_u(g) \not\subseteq V_w(g, i)\} \tag{2}$$

The exposure set $E(w)$ represents every piece of information throughout a social network uploaded by other users that contains information about w despite w's intent to keep the information private. While a single leaked friendship or wall post may pose a minimal risk to a user's privacy, we show in Section 4 how the entire exposure set can be used to infer a user's personal attributes.

An important aspect of the exposure set $E(w)$ is distinguishing information visible to the entire social network from information exposed to a limited number of users. Consider a situation where Alice posts a photo and tags Bob. If Alice allows all users $u \in U$ to view her photos and is in conflict with Bob's exposure policy, we say a *global exposure* has occurred. In this case, Bob's information is revealed to Facebook users that have no prior relationship with either Alice or Bob. Conversely, if Alice exposes Bob's information to a set of users that are friends or friends of friends, we say a *local exposure* has occurred. While Bob's information is still being revealed against his will, only users that have some pre-existing relationship with Alice can view the data, not all of Facebook.

4 Inference Techniques

While scattered details about relationships and conversations between users may not pose an obvious threat to privacy, we present two classification systems that utilize the aggregate information exposed by privacy conflicts to infer a user's sensitive attributes. These techniques highlight how seemingly innocuous data leaked by friends can be used to infer meaningful private data, illustrating the necessity of multi-party privacy in social networks. While predicting a user's personal attributes based on friends has been previously examined [17–20], we present improvements to these techniques that utilize auxiliary information including wall posts, mutual friends, and the frequency of communication between users to further refine predictions.

4.1 Threat Model

The goal of classification is to infer properties about a user based on information either intentionally revealed or unintentionally exposed due to privacy conflicts. We assume that a user restricts access to his list of friends and wall posts and that no *a priori* information about the user exists. Under this scenario, aggregating personal data requires scouring a social network for privacy conflicts that link back to the user. To accomplish this task, we assume the parties involved are marketers, political groups, and monitoring agencies [10, 11, 16] who have the resources, sophistication, and motivation to glean as much information from social networks as possible. We also assume the interested parties do not form relationships with users or their friends to circumvent privacy controls. When considering the success of gathering privacy conflicts and inferring a user's personal information, we avoid any qualitative analysis of privacy risks such as the

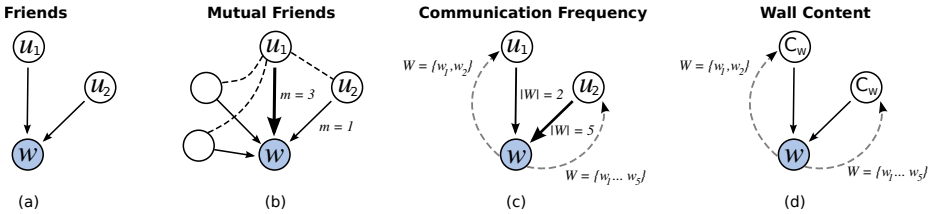


Fig. 1. Classification models for inference. Relationships and wall posts leaked by friends can be used to determine properties about the user w. These values can then be weighted based on the number of mutual friends or the frequency of communication between two friends.

damage incurred by a photo being made public. Instead, we attempt to predict eight private attributes from data exposed by privacy conflicts. Four of the attributes target personal information, including a user's gender, political views, religious views, and relationship status. The other four attributes target media interests, including a user's favorite music, movies, television shows, and books.

4.2 Analytic Techniques

In this section, we describe the development of two classifiers that take the set of information exposed about a user throughout Facebook by friends and output predictions about the user's attributes. Currently, we restrict our classifiers to analyzing leaked friend lists and wall posts. A successful prediction using leaked data means that the details exposed by friends contain enough information to further violate a user's privacy, while an unsuccessful prediction means that the leaked data was too limited to draw a meaningful conclusion about a user's attributes. When predicting personal attributes, only one prediction is correct; a user can either be liberal or conservative, but not both. Conversely, media interests represent a multi-label classification problem where users can have multiple favorite books and movies. When predicting media interests, we return up to ten predictions and evaluate whether any one of them is correct.

Baseline Classifier In order to quantify how access to auxiliary information helps to improve predictions about a user's attributes, we compare the accuracy of each classifier we develop against a baseline classifier. For each attribute, the baseline predicts the most frequent class within our data set. For multi-label attributes such as a user's favorite books where multiple predictions may be correct, the baseline returns the top ten most likely classes.

Friend Classifier Using links between friends that are publicly exposed by privacy conflicts, the friend classifier attempts to predict a user w's attributes based on other Facebook members w associates with. While a link between two users carries no explicit private data, the friend classifier builds on the assumption that if two users are friends, they likely share correlated interests. The friend classifier begins by aggregating the publicly accessible features u appearing in all of w's friends' profiles as shown in Figure 1(a). During single-label classifica-

tion, we limit the set of features aggregated to a friend's gender, political view, religious denomination, and relationship status. Multi-label classification takes a different approach, where to predict a user's musical interests, we only consider the musical interests of his friends; all other features are ignored.

Rather than naively treating each of a user's friends as being equally influential, classification attempts to distinguish between strong and weak relationships and weight features appropriately. Given a relationship (w, f) between a user w and a friend f, each feature u aggregated from f is represented as a tuple (u, m_u, w_u). The weight m_u equals the number of mutual friends shared between (w, f) that are publicly known, as shown in Figure 1(b). The goal of including m_u is to reinforce clique structures which historically share similar interests [21], while removing incidental relationships that are not part of the clique and likely to perturb classification. A similar approach is taken for communication frequency where the weight w_u is set to the number of wall messages that w has sent to f, as shown in Figure 1(c). Including w_u helps to filter out friends that rarely communicate, which was previously identified as a strong indicator of a weak relationship [22].

The resulting list of tuples (u, m_u, w_u) is binned based on distinct features and converted into a feature vector. For single-label classification, a multinomial logistic regression [23] is used to classify every user and segment the feature space into types of friends associated with a user having a specific attribute, such as being male or female. For multi-label classification where the feature space is much larger, a linear regression selects the ten most likely media interests from a user's friends exclusively, ignoring trends identified from classifying other users and their friends. Successful classification for both techniques hinges on users being biased in their selection of friends due to sharing similar interests, while unsuccessful classification would indicate a user selects friends at random.

Wall Content Classifier The wall content classifier attempts to predict a user w's personal attributes based on text recovered from w's conversations with friends. Classification begins by gathering all the wall posts written by w, but exposed to the public by w's friends. Each post is then concatenated to create a single document containing all of w's discussion that is treated as a bag of words. Using classic document classification techniques, the set of wall posts is converted into a word vector where the associated frequencies of each word are weighted using term frequency–inverse document frequency [24]. The resulting word vectors from every user are classified using a multinomial logistic regression that attempts to segment the feature space into words typically used by women rather then men, or liberals rather than conservatives. Accurate classification hinges on conversations between users differing along attribute boundaries, while inaccurate classification indicates conversations between users are homogeneous despite varying attributes among users.

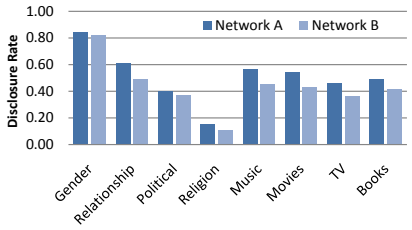


Fig. 2. Profile feature disclosure rates. Users readily supply their gender and media interests, but rarely reveal religious views.

5 Experimentation

Using the classifiers presented in Section 4, we analyze the accuracy of each technique on two real world Facebook data sets.[3] We begin by providing an overview of our data set and the frequency of potential privacy conflicts, finding that asymmetric privacy settings are common throughout Facebook. We then examine the accuracy of each classifier and whether the intuition behind each technique proved correct. Our results show classification using information gleaned from privacy conflicts consistently outperforms predictions that lack the auxiliary information, proving that conflicts can be analyzed to expose meaningful sensitive information. Further, we find that accuracy is directly related to the number of conflicts between a user and his friends. As more information is unintentionally exposed to the network, we can construct an increasingly accurate image about a user, highlighting the necessity of multi-party privacy.

5.1 Data Set

Our experimental data set consists of over 83,000 real world Facebook user profiles as shown in Table 1. The profiles are drawn from two Facebook subnetworks distinguished by geographic location, with 43,000 users associating themselves with Network A and another 40,000 users with Network B. In addition to profile pages, our data set contains over 7.5 million links between friends and 3.3 million wall posts. Of the profiles in our data set, 44% of Network A members allow a public user to view their data as opposed to 35% of Network B. This provides us with a subset of over 33,000 profiles with publicly accessible information to analyze for privacy conflicts. The rates which users reveal personal information in their profiles are shown in Figure 2. We find that users readily supply their gender (required when signing up for an account) and media interests, while less than 15% reveal a religious affiliation. After a brief preprocessing phase to correct spelling errors, group semantically similar terms, and prune unlikely labels, we identify 22 labels to describe personal attributes and over a thousand labels for media interests.

[3] It is possible – if tedious – to manually or semi-manually gather Facebook profile data without violating Facebook's Terms of Service which prohibits automated crawling.

Statistic	Network A	Network B
Profiles in data set	42,796	40,544
Fraction of Facebook subnetwork	57.70%	52.92%
Number of friends	4,353,669	3,290,740
Number of wall posts	1,898,908	1,364,691
Fraction of profiles public	44%	35%
Fraction of profiles private	56%	65%

Table 1. Our data set consists of two geographically distinct subnetworks of Facebook, amounting to over 80,000 profiles used to identify privacy conflicts and infer personal attributes.

Statistic	Network A	Network B
Number of exposed friends	1,012,280	612,387
Average exposed friends per profile	42.18	23.24
Number of exposed posts	407,278	289,877
Average exposed posts per profile	53.85	43.12

Table 2. Frequency of privacy conflicts between public and private users. An average private profile in our data set has over 80 references publicly exposed by friends with weaker privacy requirements.

5.2 Frequency of Privacy Conflicts

Analyzing our data set, we verify that asymmetric privacy requirements between friends are a common occurrence. Using each profile in our data set, we examine public lists of friends for references to private users. We repeat this same process for wall pages, identifying messages written by private users that are exposed by public pages. The results of our analysis are shown in Table 2. We identify over 1.7 million relationships and roughly 700,000 wall posts referencing private profiles that are publicly exposed by friends due to the lack of multi-party privacy controls. This amounts to approximately 96 references per user in Network A and 66 references in Network B. The skew in Network B towards fewer conflicts is a result of fewer publicly accessible pages for the network, as described earlier in Table 1. Analyzing each user's list of friends, we find on average that our data set contains information for only 35% of friends, leaving another 65% of friends with profiles that may leak private information and increase the frequency of conflicts.

5.3 Classifier Accuracy

To test the accuracy of using auxiliary information leaked by friends for predicting private attributes, we run each of the classifiers presented in Section 4 on both networks in our data set. We simulate closed profiles by concealing an open profile's attributes during classification, after which we compare the classifier's results against the true profile values. We measure the predictive success of our classifiers using standard cross-validation techniques; each classifier builds a model using 90% of the profiles in a network and is tested on the remaining 10%. This process is repeated ten times, using a different 10% of the network each round to ensure that every profile is used only once, averaging the results from each run.

Profile Attribute	# of Labels	Baseline	Friend	Wall Content
Gender	2	61.91%	67.08%	**76.29%**
Political Views	6	51.53%	**58.07%**	49.38%
Religious Views	7	75.45%	**83.52%**	53.80%
Relation Status	7	39.45%	**45.68%**	44.24%
Favorite Music	604	30.29%	**43.33%**	-
Favorite Movies	490	44.30%	**51.34%**	-
Favorite TV Shows	205	59.19%	**66.08%**	-
Favorite Books	173	42.23%	**44.23%**	-

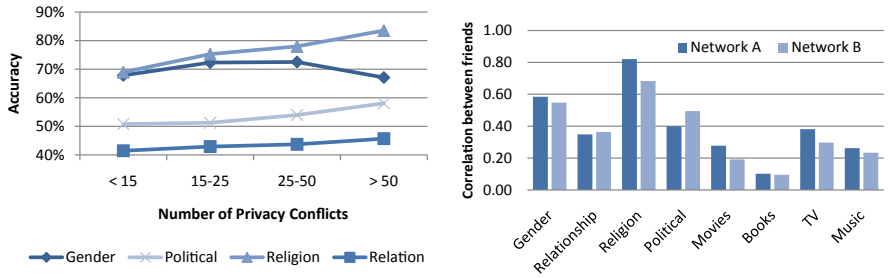
Table 3. Classifier accuracy for profiles with more than 50 privacy conflicts, representing the upper 25% of our data set. Classifiers using leaked private information consistently outperforms the baseline.

The accuracy of each classifier for profiles with over 50 privacy conflicts can be seen in Table 3. We find that the friend classifier consistently outperforms the baseline classifier, predicting profile attributes with up to 84% accuracy. Comparing the results, the wall classifier performs the best at predicting a user's gender, but fails to draw meaningful conclusions about other attributes due to the homogeneity of conversations. Accuracy for both classifiers hinges on having enough auxiliary information leaked by friends to draw meaningful predictions. Plotting accuracy as a function of privacy conflicts, we find that accuracy grows roughly linearly with the amount of exposed information, as shown in Figure 3(a). As our data set contains only 35% of potentially conflicting friends, in practice, classification will be far more accurate given a more complete data set, assuming the trend toward accuracy remains constant. We now examine each of the classifiers in detail, validating the assumptions behind each technique.

Friend Classifier The friend classifier operates on the assumption that friends have correlated features, capitalizing on information exposed by a user's friends to infer properties about the user. The friend classifier consistently outperforms the baseline, by up to 13%, for predicting a user's musical interests.

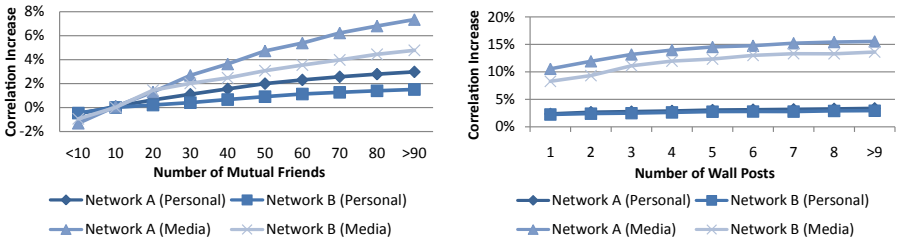
Accuracy of the friend classifier is intrinsically tied to the probability that two friends share the same feature. We measured the rates at which friends share attributes and present the results in Figure 3(b). The friend classifier can predict religion relatively well even for a limited number of samples due to the strong likelihood that two friends will share the same faith when listed. Conversely, predicting a user's gender requires far more samples to overcome the fact that most users are friends with roughly equal numbers of men and women. Surprisingly, the cross-correlation between any pair of attributes is below 20%. This means that using a friend's religion to predict a user's gender is less effective than had the friend's gender been available, but is still useful to include in classification.

To weight relationships where users are more likely to share correlated interests, the friend classifier includes information about the number of mutual friends and the frequency of communication between two users. To validate the use of both weights, we measured the correlation of attributes between two friends as a function of mutual friends, shown in Figure 4(a), and communication frequency,



(a) (b)

Fig. 3. (a) Accuracy of the friend classifier grows roughly linearly as a function of the number of privacy conflicts. (b) Correlation of attributes between two friends. Our classifiers rely on the assumption that two friends share similar interests. This is largely true with religion, but not for books.



(a) (b)

Fig. 4. (a) Analyzing the improvement of feature correlation as a function of mutual friends. Friends with large cliques of mutual friends are more likely to share features, compared to the average. (b) Analyzing the improvement of feature correlation as a function of wall posts. Friends with frequent communication tend to have stronger correlated media interests, compared to the average.

---

shown in Figure 4(b). Both figures show a tendency toward shared interests for higher numbers of mutual friends and frequent communication. To understand how these weights improve accuracy, we re-classified our data set using a friend classifier that ignored both mutual friends and wall posts. On average, including the additional weights resulted in 1-2% more accurate predictions.

Wall Classifier The wall classifier analyzes conversations leaked between friends to determine properties about a user. The results presented in Table 3 show that the classifier performs best when predicting a user's gender, but fails to produce meaningful results for all other attributes. Successful prediction of a user's gender derives from differences between the words used by women and men, while the remaining attributes such as religion or political view show no overwhelming tendency towards discussions that result in different word frequencies. Nevertheless, the appearance of terms such as sports, television shows, and news articles

all expose a users's interests and can erode privacy. We leave the application of more sophisticated document classification models for future work.

6 Enforcing Multi-Party Privacy

Having explored the extent that privacy conflicts appear throughout social networks and their potential risk, we now present a solution for enforcing multi-party privacy. Using the formalism presented in Section 3, we define a new access control framework for social network data. The framework enforces the mutual privacy requirements of all users referenced by a piece of data to prevent privacy violations, mitigating any risk of aggregating leaked information. We prototype our solution as a Facebook application that transparently enforces multi-party privacy without requiring interaction from users.

6.1 Mutual Privacy Requirements

Privacy conflicts currently arise in social networks because only the owner u of data can specify a privacy policy P_u, regardless of whether multiple users have an interest in keeping the data private. To adopt a mutually acceptable privacy policy for *all* parties, each user w referenced in content must be able to augment the policy set by u. To achieve multi-party privacy, we allow every user w to specify an exposure policy $V_w(g, i)$ for each page g and the information on that page i. The policy V_w's granularity can be page and reference specific, or alternatively, represent a policy for all pages throughout the social network. For example, a user w can specify that only w's friends can view wall posts written by w, encompassing the set of all wall pages, g, and the individual posts i. Our framework can also accommodate fine-grained policies; for example, a user w can set a policy that allows only friends and not family to view pictures posted by w's friends. In practice, we expect most users to set coarse rather than fine-grained exposure policies that restrict access to all information for a user w.

For each piece of information i on page g, the largest set of users who can view i without violating any user's privacy policy can be represented by the mutual privacy policy $P_m(g, i)$:

$$P_m(g, i) = P_u(g) \bigcap_w V_w(g, i) \tag{3}$$

P_m represents the set of users that the content owner u and all the associated parties $w \in S(i)$ mutually trust with their personal data. In the absence of mutually trusted friends, P_m tends towards the empty set, resulting in i being hidden from every user. However, the majority of the privacy conflicts we identified involve only two users, such as bidirectional links between friends, reducing the number of policies which must be satisfied. Photos and wall posts that refer to multiple users present a more complex situation where access to content is highly restricted due to multiple exposure policies. The potentially limited size of P_m is a byproduct of satisfying every user's privacy without bias; otherwise, a larger P_m would only violate one user's expectation of privacy.

For social networks that allow a user w to remove references to himself, such as with Facebook photos, multi-party privacy policies represent a stronger alternative. A user removing a reference to himself from a compromising image still leaves the privacy violating content exposed, if only harder to identify. Conversely, multi-party privacy guarantees that every user's privacy requirements are satisfied. This extends to situations where users cannot remove themselves such as with friendships, group membership, and comments, guaranteeing that privacy is always satisfied.

6.2 Prototyping Multi-Party Privacy

To demonstrate the feasibility of multi-party privacy, we create two Facebook applications that reproduce the functionality of a friend list and wall page while enforcing mutual privacy policies. These prototypes serve to show how Facebook could implement multi-party privacy; they do not replace the existing friend and wall pages which Facebook prevents from being modified by applications.

Assuming the applications are installed on a fully public profile, the privacy-enhanced friend list conceals the names of friends with exposure policies that prohibit a third party from seeing the relationship. Similarly, the privacy-enhanced wall conceals wall posts if the original sender prohibits a third party's access. Currently, if an exposure policy for a user is not specified, the application places privacy as a priority and automatically conceals references pertaining to the user. For non-public profiles where the owner policy is more restrictive than an exposure policy, the owner policy takes precedent. The result of each of these policies is a system that guarantees a user's wall posts and friends cannot be exposed against his will.

By modifying friend and wall pages to restrict access based on a reader's permissions, we are potentially changing static structures into dynamic documents that must be reprocessed each access. There is already a precedent for implementing tailored pages in Facebook, such as the news feed, which provides each user a distinct set of stories based on their interests and friends that changes as the day goes by. Enforcing multi-party privacy can thus be seen as an extension of news feeds, where the content displayed is based on privacy controls rather than interests. By adopting the enforcement of multi-party privacy, Facebook users gain control over all their private information, even if it is uploaded by another party.

7 Related Work

There is an extensive body of research on protecting and examining privacy in social networks. The most related of these works to our research are attempt to demonstrate flaws in the current privacy controls of social networks. Zheleva et al. [17] examine the risks of revealing group membership and friendships, while He et al. model correlated features between friends as a Bayesian network [18]. Adapting previous approaches to attribute inference, Mislove et al. [20] looked at community structures among friends, finding that tight-knit communities often shared highly correlated features. Our work can be seen as a refinement of their techniques, presenting new ways to identify meaningful friends and filter

relationships that are likely to impede inference. We also examine previously unexplored avenues such as wall posts for inference, pointing out that any relationship or tag between two users can potentially violate privacy.

While we limit our discussion to preventing crawling and mining by third parties, other researchers have looked at how to protect information from social network providers and server break-ins. flyByNight [25], NOYB [26], and FaceCloak [27] all use encryption or steganography to protect a user's personal information to prevent a social network operator such as Facebook from reading or mining personal data. Keys are then distributed to trusted friends out of band from the social network operator, allowing friends to decrypt profile information. Despite the potential added privacy from encryption, each of these protection mechanisms rely on the social network to keep track of friends and do not extend to content posted by friends, leaving users exposed to the inference techniques we describe.

Other research in extending social network privacy includes protecting users from third party applications. Social networks such as MySpace and Facebook allow users to install applications such as games or media plugins, in turn granting the application access to all of their personal data. Applications currently lack access control restrictions, allowing programs to offload all of a user's data in addition to that of a user's friends. Felt et al. [28] and Singh et al. [29] both propose new application architectures to restrict personal data available to applications. Because applications are granted access to both the installer's data and the installer's friend's data, application security must address the requirements of multi-party privacy to guarantee users are not put at risk by their friends.

In addition to privacy protections within social networks, data released by network operators to the public also poses a significant challenge to user privacy. De-anonymization efforts [30–33] have shown that publishing anonymized or restricted social graph information is riddled with complications. These same techniques for de-anonymization can also be used for inferring properties about data leaked by users within social networks, highlighting the need for better privacy controls that suit the range [34, 35] of each users privacy expectations.

8 Conclusion

In this paper, we have shown how existing privacy controls in social networks fail to protect a user from personal content leaked by friends. As photos, stories, and data are shared across the network, conflicting privacy requirements between friends can result in information being unintentionally exposed to the public. We formalized multi-party privacy requirements which guarantee that the privacy concerns of all users affected by an image or comment are mutually satisfied. The current lack of multi-party privacy results in scattered references to users throughout social networks that can be collected by adversaries who have the resources, sophistication, and motivation to glean as much information from social networks as possible. We have shown how seemingly innocuous references to users can be aggregated and analyzed to construct meaningful predictions about a user's personal attributes and media interests. This slow erosion of personal privacy can be prevented by the adoption of multi-party privacy controls. We

prototyped these controls for Facebook, showing how multi-party privacy can be adopted, returning control over personal data in social networks to users.

References

1. Facebook: Statistics (2009) http://www.facebook.com/press/info.php?statistics.
2. MySpace: Statistics (2009) http://www.myspace.com/statistics.
3. Miller, C.: Twitter makes itself more useful. (April 2010) http://bits.blogs.nytimes.com/2010/04/14/twitter-makes-itself-more-useful/.
4. MySpace: Privacy Policy. (2008) http://www.myspace.com/index.cfm?fuseaction=misc.privacy.
5. Facebook: Privacy Policy. (2008) http://www.facebook.com/policy.php.
6. George, A.: Living online: The end of privacy? New Scientist (September 2006)
7. Sarno, D.: Facebook founder Mark Zuckerberg responds to privacy concerns. Los Angeles Times (2009)
8. CareerBuilder: Forty-five Percent of Employers Use Social Networking Sites to Research Job Candidates, CareerBuilder Survey Finds. (2009)
9. Maternowski, K.: Campus police use Facebook. The Badger Herald (January 2006)
10. Greenberg, A.: Mining MySpace. Forbes (2007)
11. Shachtman, N.: Exclusive: U.S. Spies Buy Stake in Firm That Monitors Blogs, Tweets. Wired (2009)
12. Richmond, R.: Phishers Now Hitting Twitter. The New York Times (2008)
13. McMillan, R.: Facebook Worm Refuses to Die. PC World (2008)
14. Room, F.P.: Facebook Announces Privacy Improvements in Response to Recommendations by Canadian Privacy Commissioner . (2009)
15. Bradley, T.: Bing Lands Deals with Twitter and Facebook. PC World (2009)
16. Wright, A.: Mining the Web for Feelings, Not Facts. The New York Times (2009)
17. Zheleva, E., Getoor, L.: To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. In: Proceedings of the 18th international conference on World wide web. (2009)
18. He, J., Chu, W., Liu, Z.: Inferring privacy information from social networks. In: Intelligence and Security Informatics. (2006)
19. Becker, J., Chen, H.: Measuring Privacy Risk in Online Social Networks. Web 2.0 Security and Privacy (2009)
20. Mislove, A., Viswanath, B., Gummadi, K.P., Druschel, P.: You are who you know: Inferring user profiles in online social networks. In: Proceedings of the 3rd ACM International Conference of Web Search and Data Mining. (2010)
21. Jones, E., Gerard, H.: Foundations of social psychology. John Wiley & Sons Inc (1967)
22. Gilbert, E., Karahalios, K.: Predicting tie strength with social media. In: Proceedings of the 27th international conference on Human factors in computing systems. (2009)
23. Bohning, D.: Multinomial logistic regression algorithm. Annals of the Institute of Statistical Mathematics **44**(1) (1992) 197–200
24. Jones, K., et al.: A statistical interpretation of term specificity and its application in retrieval. Journal of documentation **60** (2004) 493–502
25. Lucas, M., Borisov, N.: flybynight: Mitigating the privacy risks of social networking. In: Proceedings of the 7th ACM workshop on Privacy in the electronic society, ACM New York, NY, USA (2008) 1–8

26. Guha, S., Tang, K., Francis, P.: NOYB: Privacy in online social networks. In: Proceedings of the first workshop on Online social networks, ACM (2008) 49–54
27. Luo, W., Xie, Q., Hengartner, U.: FaceCloak: An architecture for user privacy on social networking sites. In: Proceedings of the 2009 IEEE International Conference on Privacy, Security, Risk and Trust (PASSAT-09). (August 2009)
28. Felt, A., Evans, D.: Privacy protection for social networking APIs. 2008 Web 2.0 Security and Privacy (W2SP08) (2008)
29. Singh, K., Bhola, S., Lee, W.: xBook: Redesigning privacy control in social networking platforms. Proceedings of the 18th USENIX Security Symposium (2009)
30. Narayanan, A., Shmatikov, V.: Robust de-anonymization of large sparse datasets. In: IEEE Symposium on Security and Privacy. (2008)
31. Narayanan, A., Shmatikov, V.: De-anonymizing social networks. In: Proceedings of the IEEE Symposium on Security & Privacy. (2009)
32. Backstrom, L., Dwork, C., Kleinberg, J.: Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography. In: Proceedings of the 16th international conference on World Wide Web. (2007)
33. Bonneau, J., Anderson, J., Anderson, R., Stajano, F.: Eight friends are enough: Social graph approximation via public listings. In: Proceedings of the Second ACM EuroSys Workshop on Social Network Systems, ACM (2009)
34. Gross, R., Acquisti, A.: Information revelation and privacy in online social networks. In: Proceedings of WPES'05. (2005) 71–80
35. Acquisti, A., Gross, R.: Imagined communities: Awareness, information sharing, and privacy on the Facebook. In: Proceedings of the 6th Workshop on Privacy Enhancing Technologies (PET). (2006)