MULTI-PARTY PRIVACY RISKS IN SOCIAL NETWORKS

BY

KURT A. THOMAS

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Adviser:

Professor David M. Nicol

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 to keep 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.

We are what we pretend to be,

so we must be careful about what we pretend to be.

– Kurt Vonnegut

TABLE OF CONTENTS

CHAPTER 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, 2, 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 thesis, 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, 18, 19], 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.

CHAPTER 2

BACKGROUND AND MOTIVATION

In recent years, online social networks have exploded in popularity. Today, sites such as Facebook, MySpace, and Twitter attract nearly 500 million members combined [1, 2, 3], each allowing users to share photos, stories, and disseminate links. Implicit to the interactions within a social network is the notion of trust; users create relationships with their friends and valued media outlets, in turn receiving access to content generated by each relationship. On the heels of the widespread adoption of social networks, phishing and malware attacks have become a regular occurrence [20, 21, 22], exploiting the trust users place in their friends. At the same time, content in social networks is being used to invade user privacy [8, 6, 7], abusing the information users put online and their perception of privacy. To understand the range of threats posed against users, we motivate this thesis by exploring the importance of security and privacy before detailing what protections currently existing in social networks.

2.1 Security and Privacy

At the crux of this thesis is the desire to defend social network users against threats to their security and privacy. With the rise in success of social networks, users have become regular targets for account theft and personal information harvesting. The adversaries involved include malicious parties motivated by the monetization of stolen information [23], government surveillance [11], marketing firms, and political groups. Threats include *phishing* where the goal is to steal the account credentials of a user, allowing an adversary to masquerade as the victim or access personal details, *malware* where users are coerced into downloading malicious content by abusing a victim's trust, and *snooping* where an unwitting victim's interactions are monitored by an undesired third party.

Defending against the vast array of adversaries and threats has led to an abundance of security and privacy products that attempt to guarantee protections for systems and personal information, amounting to a multi-billion dollar industry [24]. Research in the field spans the scope of protecting industries and governments to protecting personal computers and information. Social networks present a novel area for research where users readily provide personal details meant to be shared with friends, but may unwittingly expose the information to the public. Similarly, the wealth of personal data can be used to ease coercive behaviors such as impersonating a friend or authoritative figure by abusing private information for the purpose of gaining trust. As adversaries continue to evolve their techniques for spreading spam, phishing attacks, and malware, social networks represent a largely unprotected and unstudied environment.

2.2 Social Networks

Online social networks provide a medium for individuals with similar interests to meet and share information, whether it be photos, videos, stories, or simply ideas. Popular sites such as Facebook, MySpace, Twitter, and Flickr cater to hundreds of millions of users, each providing a unique outlet to upload personal information. Due to the sensitivity of some content, social networks currently allow users to specify a range of privacy policies. The policies enacted are entirely up to a user; an individual with strong expectations of privacy can heavily restrict access to his data, while another user can divulge her content to the entire social network.

Despite the array of privacy controls available, none of the existing implementations are capable of enforcing privacy controls on data as it propagates through a social network. Similarly, there are no mechanisms to resolve conflicts between friends with incompatible privacy requirements. For instance, when a user uploads a photo and tags multiple friends in the picture, only the owner of the photo can specify a privacy policy. Every other user referenced is left without any control. While this problem exists in all online social networks, we restrict our analysis to Facebook given its status as the largest network with over 400 million users [1]. We provide a brief overview of Facebook and its privacy model before examining how users can unintentionally violate the privacy of their friends.

2.2.1 Anatomy of a Social Network Profile

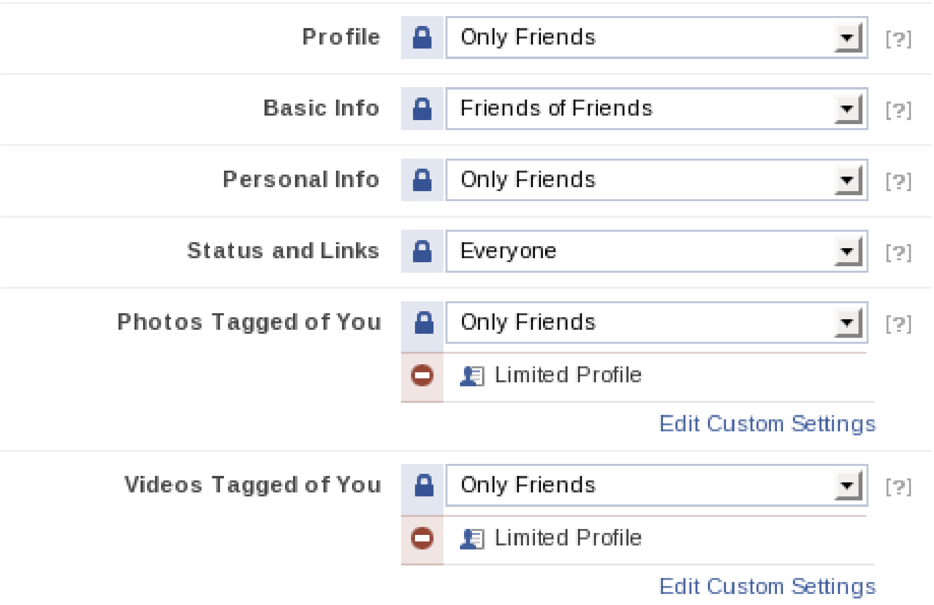
Each Facebook user is provided a *profile* consisting of a page containing *personal information*, a list of the user's *friends*, and a *wall* where friends can post comments and leave messages similar to a blog. A typical profile will contain information pertaining to the user's gender, political views, 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, Eve's friend. Each new addition is broadcasted to a user's friends, who can in turn leave comments and pass the information along to other friends.

As a user creates connections with other friends, a complex network of information begins to form. While the majority of content pertaining to a user is found on his own profile, references resulting from being tagged in a photo or mentioned in a wall post will appear on other profiles. This also includes comments posted by a user to a friend's profile or to public pages. The result is a vast collection of scattered references to a user appearing throughout a social network. Each of these references divulges some personal information to which a user should be allowed to control access.

2.2.2 Privacy Policies and Controls

In order to protect privacy, Facebook provides users with the ability to control which Facebook members can access their information. 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. A capture of Facebook's privacy controls can be seen in Figure 2.1. For each component of a user's profile, a different privacy setting can be adopted. For instance, basic information that includes a user's gender and political views can be revealed to friends of friends, while more personal information such as photos can be restricted to direct friends.

Beyond coarse-grained restrictions such as *friends* or *everyone*, users are allowed to develop



Profile: Only Friends

Basic Info: Friends of Friends

Personal Info: Only Friends

Status and Links: Everyone

Photos Tagged of You: Only Friends

⊖ Limited Profile

Videos Tagged of You: Only Friends

⊖ Limited Profile

Figure 2.1: Facebook privacy controls allow access to be restricted to friends, friends of friends, or everyone. A unique policy can be specified for each component of a user's profile.

fine-grained access control policies tuned to their personal requirements. A user can specify one subset of friends who can view photos and an alternate subset of friends who can view wall posts. These complex controls can be refined by a user to suit his expectations of privacy, but as we show next, have unseen limitations.

2.2.3 Limitations of Privacy Controls

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 or delete the message once posted. Similarly, if a user posts a photo and tags a friend, 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. When privacy conflicts expose personal information to Facebook as a whole, a user's privacy is slowly eroded. Public references can be crawled by marketers, political groups, and monitoring agencies who have the resources, sophistication, and motive to glean as much information from social networks as possible.

Similarly, undesired references can be viewed by employers [8] and law authorities [9], leaving users vulnerable to exploitation of content uploaded by friends that users never intended to be made public.

The goal of this thesis is to identify both the frequency and associated risk of privacy conflicts in Facebook. We examine how data exposed throughout Facebook can be used to infer properties about the subjects of each exposure. Our analysis highlights how leaks of seemingly innocuous data by friends can unintentionally reveal sensitive personal information, requiring a redesign of social network controls to include multi-party privacy.

CHAPTER 3

RELATED WORK

To understand the importance of social networks, we present a broad exploration of current research in online social communities. While this thesis is specifically interested in security and privacy threats within social networks, much of the existing research motivates our interest in security and privacy. As users reveal their relationships, opinions, and personal details online, the data has not only great potential for improving society and spreading information, but also equal potential for being abused for scams, unethical monitoring of individuals, and invasions of privacy.

3.1 Exploring Social Networks

The phenomenon of online social networks has generated a breadth of research interests ranging from uncovering user interaction within networks to measuring the diffusion of information among peers, characterizing the strength of relationships between network participants, and exploring how communities can be used to generate commercial and political capital. In the broadest sense, a social network is any system where a user can generate and consume content and maintain a list of social contacts [25]. This includes popular online social communities such as Facebook, MySpace, and Twitter in addition to blogging, email, instant messaging, and phone communication. Within each medium a user acts as a node within a community and makes connections with other users, exposing these relationships to the public or to a subset of the public.

3.1.1 Measuring Behavior within Networks

With the explosion in popularity of online social networks, researchers are challenged with understanding how users form relationships and whether users organize themselves into highly connected or segmented subgraphs. A study by Mislove et al. [26] found that the number relationships a user forms varies drastically across social networks, with users of the photo sharing site Flickr having on average 12 friends compared to micro-blogging Orkut users that maintained over 100 relationships. Yet for either site, it only takes roughly 5 link traversals to arrive at any other node within the social graph, replicating the results of an earlier study into what researchers called the small world experiment [27], better known as six degrees of separation.

In addition to the types of relationships formed within social networks, much of social network research has been dominated by understanding social interaction and user behavior. One study [28] found that most users spent the majority of time browsing profiles and photos of friends, during which they viewed content from both friends and friends of friends, indicating that users are interested in content generated by users they share no direct relationship with. For social networks that allow applications, there is also a strong indication that users and their friends will adopt the same applications and regularly interact [29].

Beyond the interactions of individual users, there have been a number of studies that attempt to explore socioeconomic differences among users across social networking sites, identifying motivations for joining one network over another and how behavioral patterns differ. With respect to geographic location, one study identified [30] differences across rural and urban communities in both how users interact and the number of friendships users maintained. Other work has identified how racial and socioeconomic boundaries existing in real world communities are replicated in online social networks [31] and that having existing friends within a network is a strong motivation for joining [32]. Each of these studies provides a brief glimpse into the behavior of users within networks which are constantly evolving as new features are added or as users migrate from one network to another [33].

3.1.2 Relationship Formation and Strength

One of the foundational requirements of online social networks is the ability to form relationships with other members within the community. These relationships can be with friends and family or with complete strangers; each link represents only the existence of a relationship, not the importance behind that relationship. Understanding why users form friendships has been a seminal part of researching social communities. With the advent of online social networks, understanding interactions on a wide scale has lead to a number of studies. Relationships among friends are typically noted for their *homophily* where users share strong correlations of interests and beliefs [34]. Despite paired interests, relationships fall between a spectrum of strong and weak *ties*, with strong relationships representing close friends, while weak relationships typify acquaintances [35].

To capture the spectrum of tie strength among friends, researchers have examined the formation of relationships within social networks to weight the importance of a link's existence. A study by Gilbert and Karahalios [36] developed a classification system for tie strength based on correlated educations, overlap in friend networks, similar interests, and the frequency of communication. A similar system was proposed by Xiang et al. [37] that relied on unsupervised learning that relied solely on shared connections among friends rather than overlapping features. By detecting strong relationships, advertisers and network operators can simplify the process of generating recommendations and ads, in addition to learning vital personal information about users from their friends.

3.1.3 Community Detection

While individual relationships form the basis of social networks, researchers are also particularly interested in the structures that users organize into as networks evolve. Community detection aims to identify clusters of users that are densely connected, but share few links with members outside the cluster. Segmenting a social graph into communities relies on measuring *conductance* to locate close-knit sets of users purely from links in the social graph [38, 39, 40, 41], at the same time minimizing the size of communities to fit a predefined criteria. Community membership offers important insights into shared interests among friends and how members of a group influence one another [42].

3.1.4 Influence and Information Diffusion

The rise in popularity of social networks has lead to a revolution in how information is disseminated across the internet. Whether through sharing photos, news stories, or reviews of products, social networks allow users to exert influence over other members with respect to purchasing or interest in a topic. Identifying sources of influence is a major area of study as researchers attempt to understand the effects of homophily, relationship influence, and external factors in the dissemination of data [43]. By identifying influential social network actors, marketers and political parties can more easily spread news and generate traffic [44].

Other studies of influence have targeted information diffusion of news within social networks. These include identifying popular topics across blogs [45], detecting real time events such as earthquakes by monitoring the frequency and popularity of topics [46], and mining the sentiment of comments passed between friends about commercial products [47]. Each of these techniques highlights how user interactions within social networks can prove valuable not only socially, but also commercially.

3.1.5 Recommendation Systems

As online social networks ease communication between users, they open up new avenues for marketing, politics, and managing user opinions. Advertising has been revolutionized by the use of data mining to identify products an individual is most likely to buy, which is further expanding to include recommendations from users that share similar interests and social neighborhoods [48, 49, 50]. At the core of recommendation systems is the ability to identify positive or negative inclinations to a product or user based on relationships an individual shares with other community members [51, 52]. Examples include predicting film preferences based on feedback from millions of users [53], search engine optimization, and shopping cart suggestions. Beyond pure marketing value, social networks also provide a vital resource for spreading political messages [54]. The influence and profits that can be generated from social networks make sites such as Facebook, MySpace, and Twitter prime targets for gathering personal information as well as influencing users.

3.2 Privacy in Social Networks

There is an extensive body of research on protecting and examining privacy in social networks. The topics range from identifying threats to user privacy, protecting user content, allowing responsible third-party access, and risks of releasing data sets to the public. We present an overview of each of these areas and how they impact the development of new privacy controls for social networks.

3.2.1 Elusive Privacy and Anonymity Guarantees

Provided the vast spectrum of social networking topics, this thesis is particularly interested in how the rise of social networks erodes privacy and whether existing protections are sufficient. The notion of privacy in the digital age has come into question as research redefines what data can uniquely identify an individual. In the past, researchers and governments held that personal data consisted of an individual's name, address, and other forms of unique identifiers that distinguish a person from the rest of the population. Achieving privacy or anonymity requires suppressing or obscuring this information from the public unless an individual consents to its release. These unique identifiers are defined as *personally identifiable information*, and form a foundation for research in privacy.

The challenge of privacy is to balance the desire to publicly disclose personal data, at the same time limiting its dissemination or preventing the information from being tied to any specific individual. There is a wealth of research into providing guaranteed anonymity, along with the consequences if personally identifiable data is not protected. One of the simplest guarantees for privacy is k-anonymity [55]. The goal of k-anonymity is to provide public access to content such as social network profiles that contain sensitive, private information about a user, while obscuring who the record is associated with. For instance, public access to social interactions among teens may be valuable to researchers, but release of the data must protect the identities of the individuals involved. Under k-anonymity, clues to an individual's identity are redacted or simplified until at least k other individuals are added to the data set with the exact same attributes. So long as k-anonymity is satisfied, the probability an adversary can correctly associate a record with a specific individual is $1/k$.

Developing a k-anonymous data set requires first identifying any information that can act as a *quasi identifier*, a piece of data that is available from a separate data source and potentially uniquely identifies an individual. For instance, an individual's $\{$*Zipcode, Gender, Birthday*$\}$ is enough to uniquely identify over 63% of American citizens [56], leaving a person vulnerable to degraded anonymity and privacy if the same information can be linked with data in other records such as census data, voting records, or campus and work directories. Guaranteed k-anonymity thus requires releasing records for k users with the same $\{$*Zipcode, Gender, Birthday*$\}$.

Despite providing a foundation for defining guaranteed anonymity, k-anonymity is not without its flaws. Given a set of k individuals all with identical quasi identifiers, if all of the individuals share the same attributes, then anonymity does not provide any protection. For example, consider k users in a social network who all list their political views as liberal. While an adversary cannot distinguish which record belongs to a specific user, he can easily identify that the user is liberal. The challenge of diversity among sensitive information released to the public was addressed by l-diversity [57] which states that the sensitive information disclosed under k-anonymity must have entropy greater than or equal to l. This metric was improved upon by t-closeness [58], which states that diversity should not be a metric of entropy, but that the overall distribution of sensitive attributes within a cluster of k users should match global statistical properties of users with the same quasi identifiers.

Regardless of the approach adopted to protect a user's identify, the challenge of guaranteed anonymity and privacy in the digital era is understanding what constitutes a quasi identifier. Increasingly, personal data is spread across the internet and duplicated in multiple directories with potential public access. If a single quasi identifier is overlooked, the result is the erosion of anonymity and privacy. Beyond the challenge of simply identifying quasi identifiers is the fact that as the feature space of quasi identifiers grows, the size of an anonymous grouping shrinks due to variance within identifiers [59]. This results in either prohibitive redaction of identifiers or small clusters of users which do not provide sufficient anonymity. As we explore the challenges of privacy throughout this thesis, a central theme is the discovery of new features such as relationships or communication between individuals which act as unique identifiers and leak sensitive information.

3.2.2 Threats to Privacy

With the explosion of success surrounding online social networks, researchers were quick to examine the extent of information that users expose to the public and the possible risks. A study by Gross and Acquisti [60] in 2005 of Carnegie Mellon Facebook users found that 87.8% of students disclosed their birthday, while 50% of students disclosed their political views and over 60% their favorite media. Overwhelmingly at the time, 99% of users allowed anyone at Carnegie Mellon to view their personal profile, allowing users without any relationship to view personal details and events. A subsequent survey of 294 students by Acquisti and Gross [61] found similar results for disclosure rates, going further to study each user's perception of privacy and trust. The majority of subjects in the survey were found to completely trust their direct friends, while friends of friends and Facebook itself were somewhat trusted. The study found most users are aware of Facebook's privacy controls and use them to fit their own privacy requirements, but informed data revelation remains an issue due to the complexities of online privacy.

While the media has expounded upon the risks associated with exposing personal data online [6, 7, 8, 9, 62], further research has shown that simply advocating the use of privacy controls for profile data is not enough. Advice for enacting privacy controls has largely surrounded limiting what information a user uploads and judicious use of settings to conceal personal information including home addresses, work history, and photos. However, even if a user conceals their personal attributes, Zheleva and Getoor [17] showed how revealing group membership and friends can allow an adversary to recover the concealed data through inference. Building on the assumption that users who join the same social groups or form relationships with other users implicitly share similar interests, the researchers conducted a range of data analytic techniques to predict a user's attributes based on the attributes of friends and groups revealed by a user in his profile. They found that even if a user adopts a closed profile, his privacy can be undermined if the user fails to conceal his friends and groups. Using a Facebook data set of 1500 profiles, they were able to recover a user's gender with 70% accuracy, assuming that 50% of their friends publicly disclosed their information. A similar study using synthetic data was conducted by He et al. [18] where a social network was modeled as a Bayesian network, finding that correlated features can be uncovered by examining relationships with friends.

The use of social graphs and the attributes disclosed by friends has redefined what information can be considered personally identifiable and uncovered a number of limitations of existing privacy controls. Further studies into inference attacks include a study by Becker et al. [19], where attributes such as school year, degree, hometown, and zipcode were inferred for 93 users based on their friends. Summing the frequency of each attribute among a user's friends, successful inference ranged from 96% to 19%, with attributes such as the country a user lives in being easily identifiable from friends, while a zip code or workplace was difficult to recover. Rather than adapt Facebook's privacy controls, the research made suggestions to Facebook users to either add spurious relationships to reduce the effectiveness of inference or to remove existing relationships which were likely to aid inference. The former solution requires exposing personal information to a wider audience which is anathema to privacy, while the second solution limits usability and restricts interactions within social networks.

Adapting previous approaches to attribute inference, Mislove et al. [63] looked at community structures among friends, finding that tight-knit communities often shared highly correlated features. By applying a number of varying community detection algorithms, the study was able to identify communities of users that shared the same university graduation year, department, and major. Applying the algorithms to detect arbitrary profile attributes, given a seed to begin community reconstruction, the algorithm was on average able to find 75% of users that shared the same attribute by reconstructing the community with 35% precision. The low accuracy of the algorithm was attributed to the inability to verify results for profiles that failed to list an attribute, where the algorithm marked inference as incorrect if a value was not available. Despite this defense, it remains unclear how community detection fares for arbitrary attributes compared to university attributes from data drawn from a homogeneous population.

Our work can be seen as a refinement of each these inference techniques. We go beyond examining just friends, defining a formalism that captures all possible conflicts between two users that can potentially leak information. We then present 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. It is important to note that even if a user adopts the most restrictive privacy settings, they are not protected from information exposed by friends. The work presented in this thesis attempts to rectify this privacy oversight.

3.2.3 Protecting User Data

Privacy in social networks goes beyond simply restricting access to data from other members. The potential for database break-ins and undesired snooping on the part of social network operators or untrusted third parties represents an entirely different class of threats. To mitigate these threats, flybynight [64], NOYB [65], and FaceCloak [66] employ a range of techniques to protect uploaded content from undesired third party disclosure, at the same time granting trusted friends access.

flybynight works to prevent Facebook and untrusted parties from analyzing a user's content by encrypting all data stored on Facebook servers. The flybynight application embeds itself into all of Facebook's pages using a browser plugin, extending Facebook's typical functionality. Prior to any text data being submitted to a form, embedded JavaScript performs client-side encryption using a public key system. The encrypted data can then be downloaded by a user's friends and subsequently decrypted using a private key, while unauthorized users and Facebook operators are left with unreadable encrypted data. Due to restrictions in Facebook's architecture, only messages can be encrypted; a user's social graph, images, and videos remains unprotected.

NOYB and FaceCloak provide similar protection using encryption, but obscure the presence of encryption from Facebook by replacing encrypted data with fake, but realistic looking, content. For instance, an untrusted viewer analyzing a user Bob's profile page will see *(Alice, Female, 14)* while a trusted user will see *(Bob, Male, 21).* The falsified content is randomly generated using data from public Facebook profiles, where correlated features such as *(Name, Gender)* are selected simultaneously to prevent unrealistic feature pairs. During decryption, the fake content is provided to a dictionary stored on a trusted third-party server that maps obscured content to its true value. Without a shared key between friends, the mapping cannot be reproduced.

While all three of these techniques can prevent Facebook or untrusted third parties from accessing published data, the underlying public key infrastructure remains a challenge to maintain, where trust is merely pushed from Facebook to a separate third-party server. Furthermore, encryption only applies to content such as wall posts and profile information, not the social graph of friends. Similarly, content uploaded by friends remains unprotected even if it contains references to a user who desires the content to be encrypted. These two limitations leave existing encryption defenses vulnerable to the attacks outlined in the previous section and those that we analyze in this thesis.

3.2.4 Private Third-Party Access

The complete distrust of Facebook represents only one threat model proposed by researchers. Other research in extending social network privacy includes protecting users from untrusted third party applications. Social networks such as MySpace and Facebook allow users to install applications including games and media plugins. Once installed, the applications are granted complete access to a user's personal information *and* their friend's personal information. With no access control restrictions, applications can offload all of a user's data in addition to that of hundreds of friends. To prevent this threat, Felt and Evans [67] and Singh et al. [68] both propose new application architectures to restrict personal data available to applications.

Felt and Evans prevent all access to user data from applications, instead proxying requests through the social network operator. To interact with user content, applications rely on a markup language. For instance, if the application is designed to list all of a user's friends' birthdays, the application would include markup text similar to `<display: uid=1907, birthday>`. The markup text is then parsed by Facebook before being displayed to a user, replacing the representative data with actual values.

As not all applications can operate on representational data, Singh et al. [68] present an alternative approach that employs client and server side protections to give users control over the data each application can view. For instance, a location tracking application would be granted access to a user's geographic information and GPS data, but not the user's birthday. This fine-grained compartmentalization of data requires users to acknowledge which data an application can access before installation. However, once access is granted, the user forfeits any rights on how the application data can be used.

While we do not specifically address application security in this thesis; application controls must address the requirements of multi-party privacy to guarantee users are not put at risk by their friends. The model proposed by Felt et al. translates easily to multi-party privacy, as third party applications are never provided actual data. Conversely, the compartmentalized model proposed by Singh et al. only considers the application controls of the installer, not the installer's friends. Such controls must enforce access controls mutually acceptable to all of a user's friends, as we will describe in Chapter 8.

3.2.5 Risks of Data Release

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 have shown that publishing anonymized or restricted social graph information is riddled with complications. Narayanan et al. [69, 70] showed that an anonymized graph can be deanonymized by using auxiliary information from public sources where the feature spaces of both sources overlap. In their example, Netflix released an anonymized data set of its users' movie interests. By using auxiliary data exposed by the Internet Movie Database, reviews publicly posted to the database could be matched up to the anonymized records, recovering a user's name and information.

Rather than comparing two data sets by features, a social graphs structure can be targeted. Bonneau et al. [71] examined Facebook's practice of publicly listing eight of a user's friends to search engines. By aggregating over 30,000 public listings, they showed how Facebook's social graph structure could be approximated using only a fraction of visible relationships. In contrast to earlier passive approaches of de-anonymization, Backstrom et al. [72] present a technique of injecting markers into social graphs by forming relationships with other users and creating spurious accounts. When anonymized graph data is made public, markers are re-identified and portions of the graph de-anonymized. Each new extension to de-anonymization and social graph reconstruction can in turn be used by the techniques described in this thesis, highlighting the necessity of better privacy controls.

CHAPTER 4

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 Chapter 5 where we examine how information leaked by privacy conflicts can be analyzed to infer a user's personal attributes and in Chapter 8 where we show how Facebook can be adapted to enforce multi-party privacy.

4.1 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. Currently, Facebook and other social networks recognize only one owner for content. This owner has sole control over privacy settings, leaving the concerns of other users unresolved. We specifically analyze two scenarios in Facebook — friendship and wall posts — to understand the types of information exposed by conflicts.

4.1.1 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 to publicly disclose. While Facebook provides a mechanism to conceal a user's list of friends,

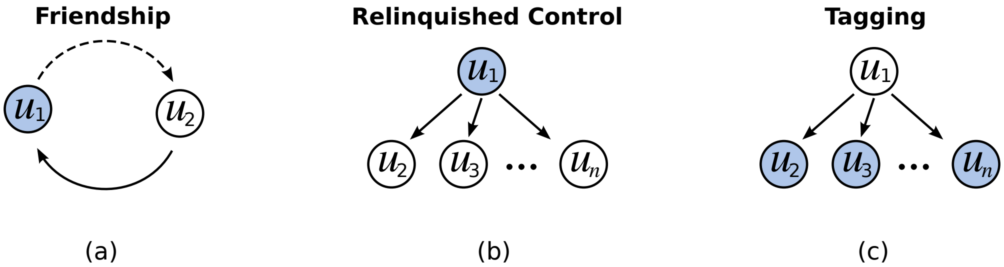


Figure 4.1: Scenarios where a user's privacy, highlighted in blue, can be violated. These include friendship, relinquishing control over content posted to other users' profiles, and being tagged in public content.

the user can only control one direction of an inherently bidirectional relationship.

Consider a scenario, depicted in Figure 4.1(a), where a user Alice adopts a policy that conceals all her friends from the public denoted by a dashed line. 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.

4.1.2 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 or any friend's wall, such as in Figure 4.1(b), 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 with tags who wish to remain anonymous, shown in Figure 4.1(c). 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.

4.1.3 Additional Conflicts

Friendship and wall posts represent only two of numerous situations where Facebook and other social networks lack multi-party privacy. Group membership, event attendance, photo tagging, video tagging, and application data all represent 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, leaving users vulnerable to content uploaded by their friends and family.

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

$$\exists i \in I : P_u(g) \nsubseteq \bigcap_{w \in S(i)} V_w(g, i) \tag{4.1}$$

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 information I on page g and no privacy conflict exists.

An important consequence of Equation 4.1 is that as the number of users referenced by a piece of information increases, in the absence of mutual friends, the union of all exposure policies tends to the empty set. This implies that for photos or wall posts referencing multiple users, its 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.

4.3 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:

$$E(w) = \{\forall (u \in U, g \in G, i \in I) : P_u(g) \nsubseteq V_w(g, i)\} \tag{4.2}$$

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 Chapter 5 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. We now explore the extent to which exposed information can be analyzed to uncover previously private properties about a user.

CHAPTER 5

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, 18, 19], 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.

5.1 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 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.

5.2 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.

5.2.1 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.

$$class(sample) = \underset{c}{\operatorname{argmax}}\ P(c) \tag{5.1}$$

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. When measuring the success of multi-label classification, we say a prediction is accurate if at least one of the ten predictions is correct. For profiles that are singleton nodes or lack friends that leak information, the baseline represents our best guess given the absence of auxiliary information.

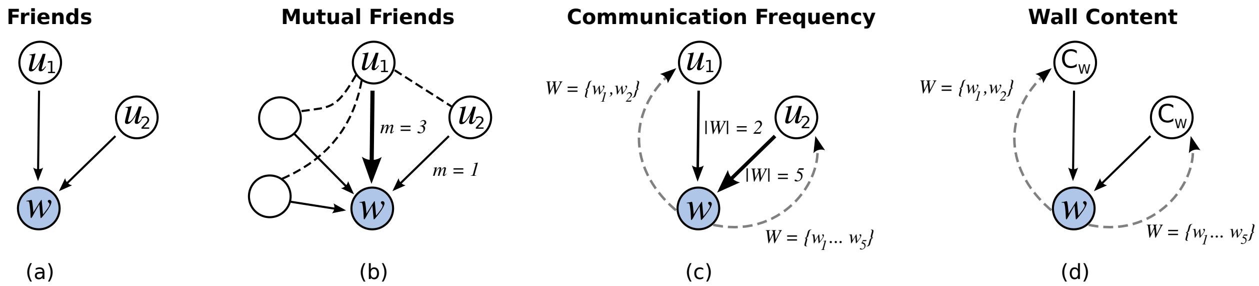


Figure 5.1: Classification models for inference. Relationships and wall posts leaked by friends can be used to determine properties about the user w. These values can then be weighted based on the number of mutual friends or the frequency of communication between two friends.

5.2.2 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 5.1(a). During single-label classification, 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 5.1(b). The goal of including m_u is to reinforce clique structures which historically share similar interests [73], 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 5.1(c). Including w_u helps to filter out friends that rarely communicate, which was previously identified as a strong indicator of a weak relationship [36].

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 [74] 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. The logistic regression treats a feature vector as a set of variables x_i, where x_1 may represent the number of male friends a user has, while x_2 represents the number of a user's friends that are Catholic. The resulting features are represented as a function:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k \tag{5.2}$$

where β_i are coefficients to the logistic regression being solved that attempt to minimize the error of the following equation for all instances in the data set:

$$f(z) = \frac{1}{1 + e^{-z}} \tag{5.3}$$

The resulting solution $f(z)$ is the probability that the relation being tested is true, yielding only a binary answer for an instance being a member of a class. To extend this approach to allow testing multiple classes where more than a binary answer is required, such as predicting a user's political views, the logistic regression is extended to a multinomial logistic regression.

For each class in $j = 0 \cdots n$, Equation 5.2 is modified such that each class has its own unique regression coefficients:

$$z = \beta_{(0,j)} + \beta_{(1,j)} x_1 + \beta_{(2,j)} x_2 + \beta_{(3,j)} x_3 + \cdots + \beta_{(k,j)} x_k \tag{5.4}$$

To determine which of the possible labels is most likely for an instance, the following equation is solved for each $j = 1 \cdots n$, while $j = 0$ acts as a base case:

$$\Pr(c = j) = \frac{\exp(X\beta_j)}{1 + \sum_{j=1}^{J} \exp(X\beta_j)} \tag{5.5}$$

$$\Pr(c = 0) = \frac{1}{1 + \sum_{j=1}^{J} \exp(X\beta_j)} \tag{5.6}$$

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.

5.2.3 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 [75].

Term frequency is computed by finding how often a term t_i appears in a particular set of wall posts j, compared to all other terms t in the wall posts:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{5.7}$$

A similar process is conducted to determine inverse document frequency, where the number of wall posts w containing a particular term t_i is compared against the total number of wall posts W:

$$idf_i = log\frac{|W|}{|\{w : t_i \in w\}|} \tag{5.8}$$

The final term frequency–inverse document frequency is computed as:

$$tf\_idf_{i,j} = tf_{i,j} \times idf_i \tag{5.9}$$

The use of term frequency–inverse document frequency allows words that appear only for certain classes to act as strong weights, while terms used frequently regardless of a user's attributes are ignored. 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.

CHAPTER 6

EXPERIMENTATION

Using the classifiers presented in Chapter 5, we analyze the accuracy of each technique on two real world Facebook data sets.[1] 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.

6.1 Data Set

For our analysis, we obtained a data set from two distinct geographic subnetworks of Facebook, each consisting of over 70,000 users. The data set consists of randomly sampled Facebook IDs drawn from a set of profiles guaranteed to be part of the same subnetwork. By relying on a random sample, our data set consists of both highly connected profiles as well as singleton nodes and unconnected subgraphs which would otherwise be impossible to encounter from relying on a depth or breadth first search. Overall, our sample consists of over 50% of each network.

The resulting experimental data set consists of over 83,000 real world Facebook user profiles as shown in Table 6.1. 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

[1] 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.

Table 6.1: 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**
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%

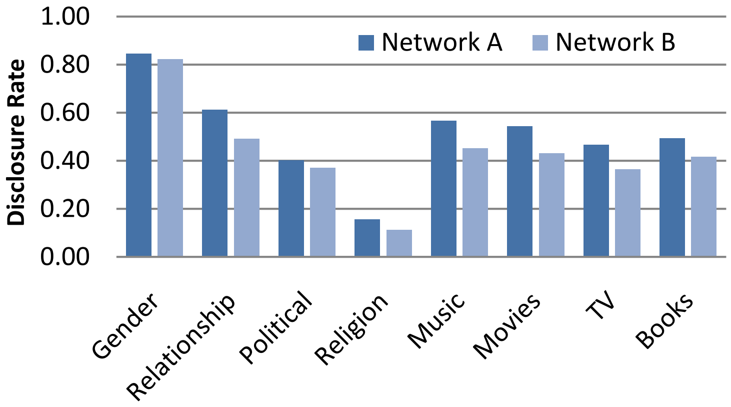


Figure 6.1: Profile feature disclosure rates. Users readily supply their gender and media interests, but rarely reveal religious views.

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 6.1. 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.

Prior to using the data, we run each recovered profile through a brief pre-processing stage. We manually developed filters for each network to group semantically similar terms listed as religions and political views, reducing the total number of labels to simplify classification. A similar approach is taken with media attributes, where each term is run through a search engine and encyclopedia to group titles all originating from the same series, correct spelling mistakes, and

Table 6.2: 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.

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

resolve ambiguities surrounding abbreviations. The resulting labels for both attribute types are pruned based on likelihood, resulting in 22 labels to describe personal attributes and over a thousand labels for media interests.

6.2 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 6.2. 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 6.1. 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.

6.3 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 Chapter 5 on both networks in our data set.

Table 6.3: 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.

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%**	-

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.

The accuracy of each classifier for profiles with over 50 privacy conflicts can be seen in Table 6.3. 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 6.2. 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.

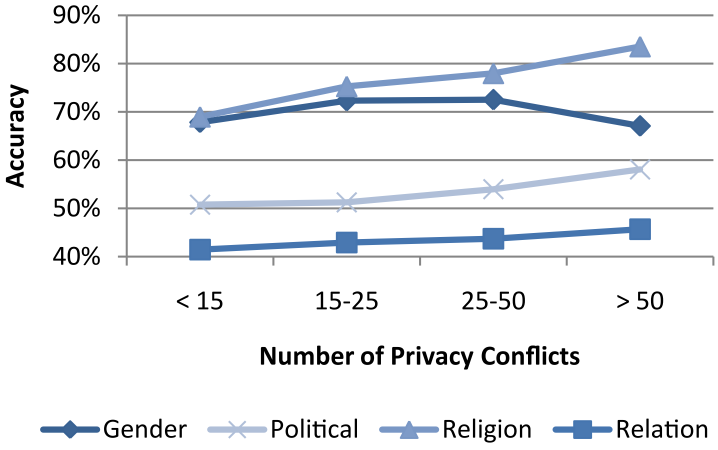


Figure 6.2: Accuracy of the friend classifier grows roughly linearly as a function of the number of privacy conflicts.

6.3.1 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 for 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 6.3. 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 most users are friends with roughly equal numbers of men and women.

To understand the cross-correlation of features, we took all relationships where two users shared the same attribute and identified whether a second, distinct attribute was more likely to match. Surprisingly, the cross-correlation between any pair of attributes is below 20%, as shown in Table 6.4. 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

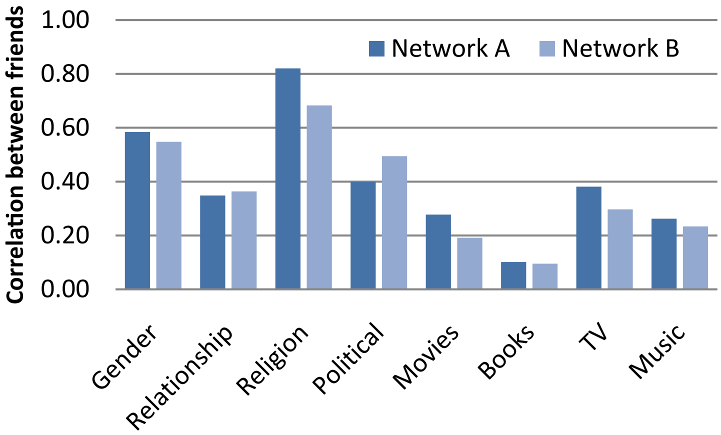


Figure 6.3: Correlation of attributes between two friends. Our classifiers rely on the assumption that two friends share similar interests. This is largely true for religion, but not for books.

Table 6.4: Cross-correlation between attributes of two friends. Most non-media attributes show limited to no correlation, while media interests show stronger correlation.

	Gender	**Relation**	**Religion**	**Political**	**Movies**	**Books**	**TV**	**Music**
Gender	1.00	0.04	-0.01	-0.02	0.05	0.03	0.06	0.02
Relation	0.04	1.00	-0.05	-0.02	0.02	0.01	0.03	0.02
Religion	-0.01	-0.05	1.00	0.01	0.00	0.08	-0.05	-0.02
Political	-0.02	-0.02	0.01	1.00	0.02	0.02	0.01	0.04
Movies	0.05	0.02	0.00	0.02	1.00	0.17	0.18	0.20
Books	0.03	0.01	0.08	0.02	0.17	1.00	0.12	0.11
TV	0.06	0.03	-0.05	0.01	0.18	0.12	1.00	0.13
Music	0.02	0.02	-0.02	0.04	0.20	0.11	0.13	1.00

correlation of attributes between two friends as a function of mutual friends, shown in Figure 6.4, and communication frequency, shown in Figure 6.5. Both figures show a tendency towards 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, with only religion showing an unexpected drop in accuracy.

6.3.2 Wall Classifier

The wall classifier analyzes conversations leaked between friends to determine properties about a user. The results presented in Table 6.3 show that the classifier performs best when predicting a

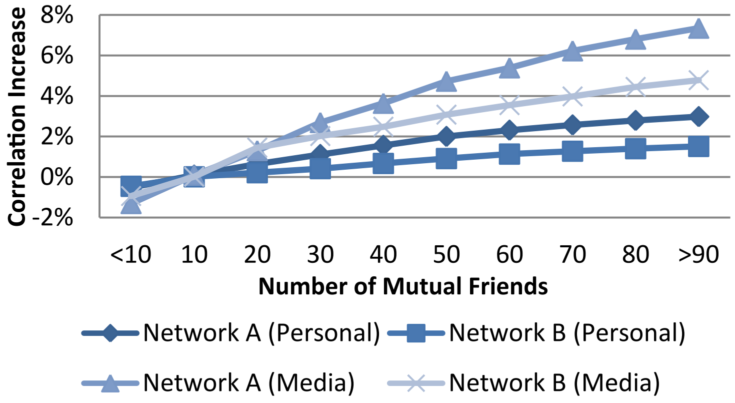


Figure 6.4: Analyzing the improvement of feature correlation as a function of mutual friends. Friends with large cliques of mutual friends are more likely to share features, compared to the average.

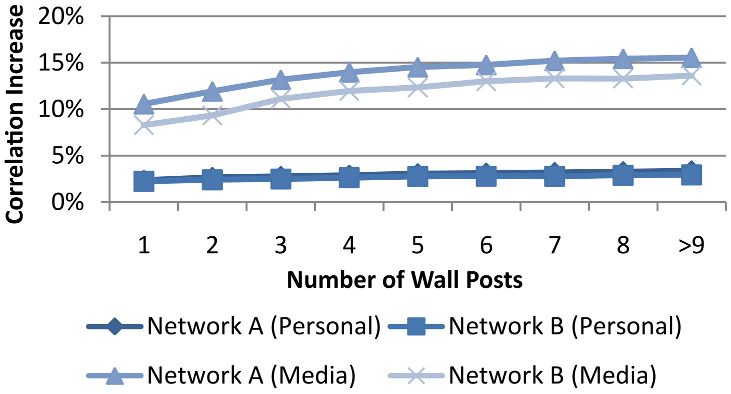


Figure 6.5: Analyzing the improvement of feature correlation as a function of wall posts. Friends with frequent communication tend to have stronger correlated media interests, compared to the average.

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 toward 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 application of more sophisticated document classification models for future work.

CHAPTER 7

SIMPLIFYING INFERENCE WITH SAMPLING

The primary challenge for an adversary in inferring sensitive attributes about a user is to obtain enough auxiliary information from friends and family to reconstruct an accurate image of a user's behavior. While the ideal scenario for an adversary would be to have all publicly available Facebook content, the sheer size of the online social network and the data generated each day make it impractical for most crawlers to keep pace with over 400 million users. In this chapter, we examine the effectiveness of random sampling for predicting attributes among a user's set of friends, finding that sampling with a low error rate is largely unachievable due to the rarity of conflicts and attribute disclosure. To remedy this fact, we present a stratified sampling approach that exploits bias in network membership to discover conflicts more easily than a random crawl.

7.1 Sampling Overview

The goal of random sampling is to make a prediction about an overall population based on a sample, or subgroup, of the population. Specifically, we are interested in determining statistics about a population of friends, such as how many friends are male or liberal, based on a small sample of those friends. For these cases, the *proportion* of users possessing a specific attribute can be modeled as a Bernoulli distribution with mean p and variance $p(1-p)$. Using classical sampling theory, we can determine a sample mean, construct a confidence interval, and calculate the sample size required to guarantee an upper bound on the error of an estimate.

Given a sample of n subjects from a total population N, the sample value y_i can be determined using the following relationship:

$$y_i = \begin{cases} 1 & \text{if a desired attribute is present, i.e. male} \\ 0 & \text{if a desired attribute is not present, i.e. not male} \end{cases}$$

The sample proportion $\hat{p}$ and variance s^2 for n samples can be computed as:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{n}{n-1}\hat{p}(1-\hat{p})$$

while an unbiased estimator of the variance of $\hat{p}$ is:

$$v\hat{a}r(\hat{p}) = \frac{s^2}{n} = \frac{\hat{p}(1-\hat{p})}{n-1}$$

For a given sample, it is important to determine how well the prediction $\hat{p}$ of a sample's proportion represents the true population proportion p. A confidence interval can be calculated for $\hat{p}$ assuming a normal distribution:

$$\hat{p} \pm z\sqrt{v\hat{a}r(\hat{p})}$$

where z represents the upper $\alpha/2$ point of a normal distribution and α determines the *confidence level* $1-\alpha$, typically 95% ($z = 1.96$). Using this bound, 95% of samples of size n will be within $\pm e$ of the true proportion p. An upper bound on the error e between the population proportion p and the estimate $\hat{p}$ for a given sample size n can then be determined from:

$$e^2 = \frac{z^2 v\hat{a}r(\hat{p})}{n} = \frac{z^2\hat{p}(1-\hat{p})}{n-1} \tag{7.1}$$

For any proportion, variance is bounded between $(0, 0.25)$ as the Bernoulli distribution bounds p between $(0, 1)$. To determine an upper bound for an estimator's error, the worst case variance $\sigma^2 = .25$ can be assumed, reducing the error bound to only a function of n.

Rather than finding the error of an estimator for an arbitrary sample, an upper bound on error can be initially specified which determines the minimum sample size required to achieve the error bound. Rearranging Equation (7.1), n can be computed with the following equation:

$$n = \frac{z^2 v\hat{a}r(\hat{p})}{e^2}$$

By estimating $v\hat{a}r(\hat{p}) = max(\sigma^2) = .25$, only e is left to be specified before gathering a sample.

Table 7.1: Sample size as a function of error bound at 95% confidence, assuming the worst case variance. Less stringent requirements on error reduce the number of samples required.

Error	**Sample Size**
3%	1068
5%	385
10%	97
15%	43
20%	25
30%	11

7.2 Sampling a Set of Friends

Using classical sampling, we aim to determine how many relationships must be recovered to draw a meaningful conclusion about the entire set of a user's friends, simplifying inference. Given a set of N friends, a single sample y_i must meet two criteria to be used to estimate properties about N. First, the sample must have privacy settings that leak auxiliary information about a user; without privacy conflicts, no useful data can be recovered. This is analogous to a response rate in classical sampling where all survey takers will not respond with meaningful results. Secondly, a sample's profile must contain attributes that can be used for inference. If a sample profile reveals a relationship, but does not contain a gender, political view, or similar attribute, knowing the relationship does not erode privacy.

Assuming an infinite population of friends for a user, satisfying a confidence level of 95% with an error bound of 3% requires 1068 usable profiles. A weaker requirement on the error bound will require fewer samples, as shown in in Table 7.1. In practice however, a user will not have an infinite number of friends, requiring an estimate for sample size to be modulated using a *finite population correction factor*:

$$fpc = \left(\frac{N-n}{N-1}\right)$$

Use of the correction factor results in an overall sample size of n_1, where n_0 denotes the original sample size prior to correction, drawn from Table 7.1:

$$n_1 = \frac{1}{(N-1)/Nn_0 + 1/N}$$

Table 7.2: Sample size as a function of error assuming a population $N = 200$, comparing ideal and non-ideal situations. A sample size of - indicates the desired error bound is impossible to achieve given the expected number of unusable samples.

Error	**Sample Size** ($r \times d_a = 1$)	**Sample Size** ($r \times d_a = 0.2$)
3%	169	-
5%	132	-
10%	66	-
15%	36	177
20%	22	108
30%	11	51

When formulating both n_0 and n_1 we assumed an ideal situation where all profile samples contained leaked data and personal attributes. In fact, finding profiles from the set of a user's friends that meet both criteria is not a trivial task. Using our Facebook data set, we measured both the conflict rate r and the disclosure rate for each attribute a, denoted d_a. On average, d_a will vary across attributes with a range of 18% to 84%, while r is roughly 40% with slight variations within Facebook subnetworks. Assuming there are no external queues to distinguish samples that meet both criteria, we must modify our earlier equations for sample size to correct for samples drawn that turn out to be invalid. Finding n valid samples can be viewed as a negative binomial distribution with the probability of success equal to $r \times d_a$. Under these conditions, the expected value of n is equal to:

$$n = n_1/(r \times d_a)$$

To understand the effects of response and disclosure rate on drawing a random sample, we compare the required sample size to achieve varying error bounds in both an ideal and realistic scenario, shown in Table 7.2. Assuming that a user has $N = 200$ friends and a 10% error bound is required, at least 66 friends must be sampled. Conversely, achieving such an error bound when $r \times d_a = 0.2$ is impossible; too many friends will have concealed attributes and not leaked auxiliary information, capping the potential accuracy of any sample set. These results show that achieving a low error bound requires sampling all of a user's friends, negating any benefit that sampling might provide.

Table 7.3: Average number of relationships and wall posts found in public profiles within our data set, separated by geographic network.

Statistic	Network A	Network B
Average number of relationships	232	232
Standard deviation	247	252
Average number of wall posts	100	97
Standard deviation	84	85

7.3 Complexities of Generating a Sample

The final challenge of drawing a sample from a user's friends is that social graph information and the location of wall posts are not known *a priori*. Instead, relationships and wall posts pertaining to a specific user must be recovered by crawling an entire social network in search of privacy conflicts. For a single user being targeted by inference, finding the user's friends out of 400 million Facebook members is akin to finding a needle in a haystack. Despite the overwhelming number of users, we explore a number of techniques and observations that reduce the search space for conflicts.

7.3.1 Network Diversity

An average public profile within our data set will contain 232 friends and 98 wall posts, with more detailed statistics available in Table 7.3. As such, a random user drawn from Facebook's entire population has an exceedingly small likelihood of containing sensitive information pertaining to a second, distinct user. This search space can be reduced if a user's hometown, university, or workplace is known, allowing a crawler to search only a subnetwork rather than all of Facebook. Examining the interaction of users within Facebook, we present a histogram of the diversity of subnetwork behavior among a user's friends in Figure 7.1. On average, a user will form relationships with 35 distinct subnetworks. The large fraction of users containing friends from less than 4 subnetworks results from 27% of users within our data set having fewer than 10 friends, reducing the likelihood of multiple network associations.

The diversity of subnetworks associated with a user's friends represents the diffusion of relationships as friends spread across universities, new hometowns, and workplaces, as well as the formation of random relationships that exist entirely online. Despite this diffusion, on average 42% of a user's friends will be from the same subnetwork, as shown in Figure 7.2. Assuming similar levels of friend concentration persist across other Facebook subnetworks, gathering data about a particular user should begin by initially targeting the networks they are most likely involved in.

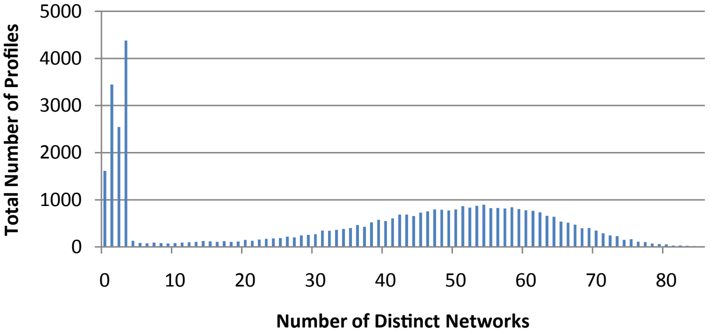


Figure 7.1: Diversity of network membership among friends. On average, a user's friends will originate from over 35 distinct subnetworks.

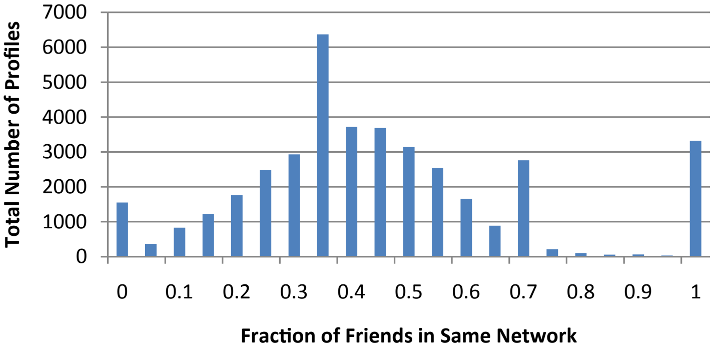


Figure 7.2: Fraction of friends within the same subnetwork for Facebook dataset. On average, 42% of a user's friends will be in the same subnetwork.

7.3.2 Privacy Conflict Recovery Rates

Rather than targeting a specific Facebook user, marketers and political organizations may be interested in targeting a specific population. As Facebook already provides clear boundaries for network membership, we can again exploit the concentration of relationships within a specific subnetwork to identify privacy conflicts and improve inference. Assuming that a random crawl is performed within a subnetwork, we measure the rate at which conflicts are recovered for *all* users within the network. As shown in Figure 7.3, on average 1,400 profiles must be crawled to recover one conflict for over 50% of users within the subnetwork. These statistics include private profiles which do not leak relationship information. If 40% of profiles are public, then we expect 560 public profiles must be crawled to recover at least one conflict for the majority of users. As discussed previously, only 40% of relationships recovered will point to another user within the subnetwork; all other relationships will reference a separate subnetwork that is not targeted for analysis. The same crawling approach was repeated for wall pages, with the results shown in Figure 7.4. Due to a profile containing fewer wall posts than relationships, a crawl requires on average 5,300 profiles to recover one wall post for over 50% of users.

While ideally all conflicting wall posts and relationships can be recovered by a complete crawl, stratified network mining represents a simple solution to recovering a majority of conflicts. Once a primary network has been crawled, random sampling of public profiles can determine the next subnetwork with the highest concentration of conflicts, with the process repeating until an acceptable number of conflicts are recovered. One consequence of stratified crawling is bias within subnetworks. For instance, the concentration of liberals or conservatives may vary vastly by network, biasing a sample towards one extreme. The development of an unbiased crawler is left for future work.

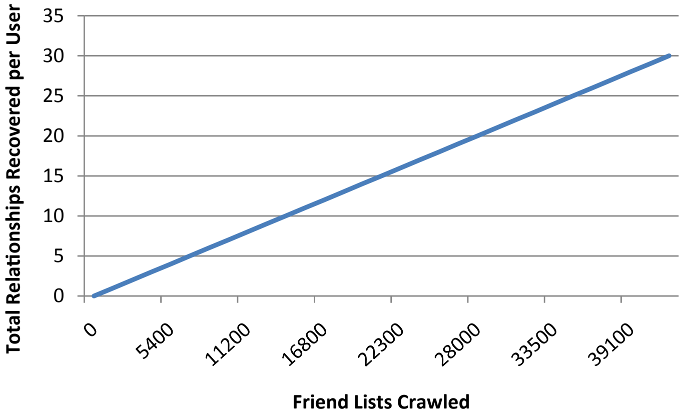


Figure 7.3: Median number of relationships recovered due to privacy conflicts as a function of friend lists crawled. On average, every 1,400 profiles crawled reveal 1 relationship for each of 40,000 users.

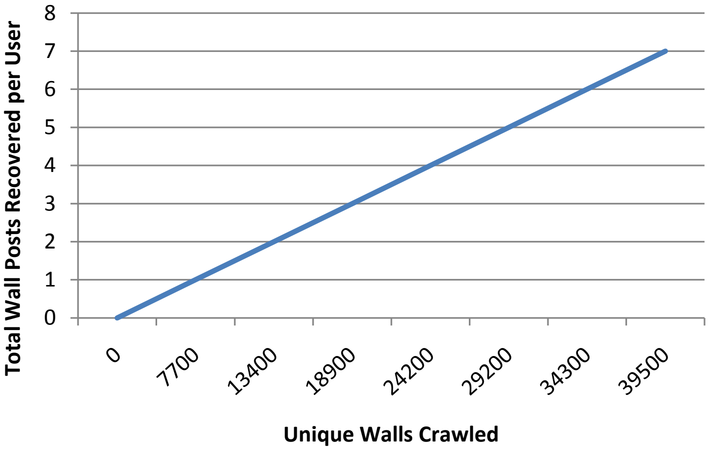


Figure 7.4: Median number of wall posts recovered due to privacy conflicts as a function of wall pages crawled. On average, every 5,300 profiles crawled reveal 1 wall post for each of 40,000 users.

CHAPTER 8

ENFORCING MULTI-PARTY PRIVACY

Having explored the extent of privacy conflicts throughout social networks and their potential risk, we now present a solution for enforcing multi-party privacy. Using the formalism presented in Chapter 4, 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.

8.1 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 or tagged 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 post 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 $P_m(g,i)$:

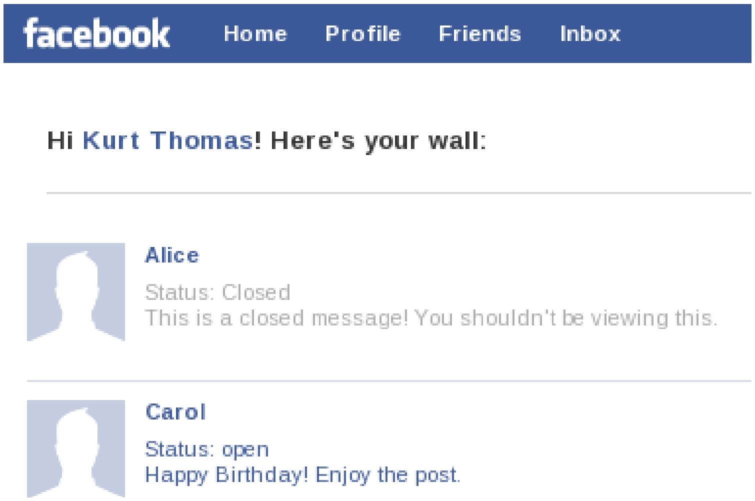
$$P_m(g,i) = P_u(g) \bigcap_w V_w(g,i) \tag{8.1}$$

P_m represents the set of users that the content owner u and all the associated parties $w \in W$ mutually trust with their personal data. In the absence of mutually trusted friends, $P_m \to \emptyset$, meaning that i will be 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 tagging multiple users present a more complex situation where access to content is highly restricted. 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 and untagging photos, multi-party privacy policies represent a stronger alternative. A user untagging 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 untag themselves such as with friendships, group membership, and comments, guaranteeing that privacy is always satisfied.

8.2 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. Because the access policy for each third party varies, the data displayed on a page is uniquely tailored on a per user basis.



Hi Kurt Thomas! Here's your wall:

Alice

Status: Closed
This is a closed message! You shouldn't be viewing this.

Carol

Status: open
Happy Birthday! Enjoy the post.

Figure 8.1: Screenshot of the privacy aware wall. Wall posts by users with private exposure policies are grayed out to simulate their removal.

Determining whether a third party can view a relationship or wall post is based on each user's owner and exposure policy. Facebook's application platform currently does not provide information on a user's privacy policy such as restricting access to *friends* or *friends of friends*. To overcome this limitation, we probe a user's privacy policy by attempting to view the user's data from the vantage point of a crawler lacking any relationships with other Facebook members. Using this method, we can discern whether a user's profile is public or private, but not any of the granularity in between. If a user u allows public access, we set their owner policy equal to their exposure policy and say that any Facebook user can view u's information including content that references u. Similarly, if a user u has a private profile, all references to u in wall posts and friend lists are removed. This technique prevents a crawler from gleaning any information about u due to references made by friends, regardless of the privacy policies adopted by u's friends. Equally important, the privacy controls we have presented are completely transparent to users. Once an exposure policy is set, the server performs all enforcement without requiring user action. A capture of the privacy enhanced wall application can be seen in Figure 8.1. Despite the wall being public, entries by users with exposure policies that restrict public access (or that cannot be determined) are grayed out to represent removal, while non-conflicting entries are displayed as usual.

By modifying friend and wall pages to restrict access based on the 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, mainly the news feed, which provides each user a distinct set of stories based on their interests and friends, changing readily as the day goes by. The enforcement of 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.

CHAPTER 9

CONCLUSION

In this thesis, 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 motive 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 gender, political views, religion, and relationship status, in addition to a user's favorite music, movies, television shows, and books. 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] E. Schonfeld, “Twitter reaches 44.5 million people worldwide in June (comScore),” *TechCrunch*, 2009. [Online]. Available: http://techcrunch.com/2009/08/03/twitter-reaches-445-million-people-worldwide-in-june-comscore/

[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] A. George, “Living online: The end of privacy?” *New Scientist*, September 2006. [Online]. Available: http://www.newscientist.com/article/mg19125691.700-living-online-the-end-of-privacy.html

[7] D. Sarno, “Facebook founder Mark Zuckerberg responds to privacy concerns,” *Los Angeles Times*, 2009. [Online]. Available: http://latimesblogs.latimes.com/technology/2009/02/facebook-founde.html

[8] CareerBuilder, “Forty-five percent of employers use social networking sites to research job candidates, CareerBuilder survey finds,” 2009. [Online]. Available: http://www.careerbuilder.com/share/aboutus/pressreleasesdetail.aspx?id=pr519&sd=8%2F19%2F2009&ed=12%2F31%2F2009

[9] K. Maternowski, “Campus police use Facebook,” *The Badger Herald*, January 2006. [Online]. Available: http://badgerherald.com/news/2006/01/25/campus_police_use_fa.php

[10] A. Greenberg, “Mining MySpace,” *Forbes*, 2007. [Online]. Available: http://www.forbes.com/2007/08/02/myspace-privacy-data_mining-tech-cx-ag-0802myspace.html

[11] N. Shachtman, “Exclusive: U.S. spies buy stake in firm that monitors blogs, tweets,” *Wired*, 2009. [Online]. Available: http://www.wired.com/dangerroom/2009/10/exclusive-us-spies-buy-stake-in-twitter-blog-monitoring-firm/

[12] R. Richmond, “Phishers now hitting Twitter,” *The New York Times*, 2008. [Online]. Available: http://gadgetwise.blogs.nytimes.com/2009/05/26/phishers-now-hitting-twitter/

[13] R. McMillan, “Facebook worm refuses to die,” *PC World*, 2008. [Online]. Available: http://www.pcworld.com/article/155039/facebook_worm_refuses_to_die.html

[14] Facebook, “Facebook announces privacy improvements in response to recommendations by Canadian privacy commissioner,” 2009. [Online]. Available: http://www.facebook.com/press/releases.php?p=118816

[15] T. Bradley, “Bing lands deals with Twitter and Facebook,” *PC World*, 2009. [Online]. Available: http://www.pcworld.com/businesscenter/article/174076/bing_lands_deals_with_twitter_and_facebook.html

[16] A. Wright, “Mining the Web for feelings, not facts,” *The New York Times*, 2009. [Online]. Available: http://www.nytimes.com/2009/08/24/technology/internet/24emotion.html

[17] E. Zheleva and L. Getoor, “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, pp. 531–540.

[18] J. He, W. Chu, and Z. Liu, “Inferring privacy information from social networks,” in *Proceedings of IEEE International Conference on Intelligence and Security Informatics*, 2006, pp. 154–165.

[19] J. Becker and H. Chen, “Measuring privacy risk in online social networks,” *Web 2.0 Security and Privacy*, 2009. [Online]. Available: http://w2spconf.com/2009/papers/s2p2.pdf

[20] E. Mills, “Twitter hit with second phishing attack this week,” *CNET News*, 2010. [Online]. Available: http://news.cnet.com/8301-27080_3-10459108-245.html

[21] E. Mills, “Facebook hit by phishing attacks for a second day,” *CNET News*, 2009. [Online]. Available: http://news.cnet.com/8301-1009_3-10230980-83.html

[22] D. Ionescu, “Twitter warns of new phishing scam,” *PCWorld*, 2009. [Online]. Available: http://www.pcworld.com/article/174660/twitter_warns_of_new_phishing_scam.html

[23] J. Franklin, V. Paxson, A. Perrig, and S. Savage, “An inquiry into the nature and causes of the wealth of Internet miscreants,” in *Proceedings of ACM Conference on Computer and Communications Security*, October 2007, pp. 375–388.

[24] G. Ratnam, “Lockheed, Boeing tap $11 billion cybersecurity market,” *Bloomberg*, 2008. [Online]. Available: http://www.bloomberg.com/apps/news?pid=20601204&sid=an2_Z6u1JPGw

[25] D. Boyd and N. Ellison, “Social network sites: Definition, history, and scholarship,” *Journal of Computer Mediated Communication*, vol. 13, no. 1, p. 210, 2007.

[26] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 2007, pp. 29–42.

[27] J. Travers and S. Milgram, “An experimental study of the small world problem,” *Sociometry*, vol. 32, no. 4, pp. 425–443, 1969.

[28] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, “Characterizing user behavior in online social networks,” in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference*, 2009, pp. 49–62.

[29] A. Nazir, S. Raza, and C. Chuah, "Unveiling Facebook: A measurement study of social network based applications," in *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, 2008, pp. 43–56.

[30] E. Gilbert, K. Karahalios, and C. Sandvig, "The network in the garden: An empirical analysis of social media in rural life," in *Proceeding of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 1603–1612.

[31] D. Boyd, "The not-so-hidden politics of class online," Microsoft Research, Redmond, WA, Tech. Rep. MSR-TR-2009-2018, 2009.

[32] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: Membership, growth, and evolution," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, pp. 44–54.

[33] M. Torkjazi, R. Rejaie, and W. Willinger, "Hot today, gone tomorrow: On the migration of MySpace users," in *Proceedings of the 2nd ACM Workshop on Online Social Networks*, 2009, pp. 43–48.

[34] M. McPherson, L. Smith-Lovin, and J. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[35] M. Granovetter, "The strength of weak ties: A network theory revisited," *Sociological Theory*, vol. 1, pp. 201–233, 1983.

[36] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, 2009, pp. 211–220.

[37] R. Xiang, J. Neville, and M. Rogati, "Modeling relationship strength in online social networks," *to appear in Proceedings of the 19th International Conference on World Wide Web*, 2010. [Online]. Available: www.cs.purdue.edu/homes/neville/papers/xiang-neville-www2010.pdf

[38] A. Clauset, M. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, no. 6, p. 66111, 2004.

[39] R. Andersen and K. Lang, "Communities from seed sets," in *Proceedings of the 15th International Conference on World Wide Web*, 2006, pp. 223–232.

[40] F. Luo, J. Wang, and E. Promislow, "Exploring local community structures in large networks," in *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, 2008, pp. 387–400.

[41] M. Newman, "Detecting community structure in networks," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 321–330, 2004.

[42] Y. Matsuo and H. Yamamoto, "Community gravity: Measuring bidirectional effects by trust and rating on online social networks," in *Proceedings of the 18th International Conference on World Wide Web*, 2009, pp. 751–760.

[43] A. Anagnostopoulos, R. Kumar, and M. Mahdian, “Influence and correlation in social networks,” in *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, pp. 7–15.

[44] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 137–146.

[45] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, “Information diffusion through blogspace,” in *Proceedings of the 13th International Conference on World Wide Web*, 2004, pp. 491–501.

[46] T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake shakes Twitter users: Real-time event detection by social sensors,” *to appear in Proceedings of the 19th International Conference on World Wide Web*, 2010. [Online]. Available: http://ymatsuo.com/papers/www2010.pdf

[47] B. Jansen, M. Zhang, K. Sobel, and A. Chowdury, “Twitter power: Tweets as electronic word of mouth,” *Journal of the American Society for Information Science and Technology*, vol. 60, no. 11, pp. 2169–2188, 2009.

[48] H. Kautz, B. Selman, and M. Shah, “Referral Web: Combining social networks and collaborative filtering,” *Communications of the ACM*, vol. 40, no. 3, pp. 63–65, 1997.

[49] P. Resnick and H. Varian, “Recommender systems,” *Communications of the ACM*, vol. 40, no. 3, p. 58, 1997.

[50] J. Breese, D. Heckerman, C. Kadie et al., “Empirical analysis of predictive algorithms for collaborative filtering,” in *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 1998, pp. 43–52.

[51] C. Forman, A. Ghose, and B. Wiesenfeld, “Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets,” *Information Systems Research*, vol. 19, no. 3, pp. 291–313, 2008.

[52] J. Leskovec, D. Huttenlocher, and J. Kleinberg, “Predicting positive and negative links in online social networks,” *to appear in Proceedings of the 19th International Conference on World Wide Web*, 2010. [Online]. Available: http://www.cs.cornell.edu/Info/People/kleinber/www10-signed.pdf

[53] R. Bell and Y. Koren, “Lessons from the Netflix prize challenge,” *ACM SIGKDD Explorations Newsletter*, vol. 9, no. 2, pp. 75–79, 2007.

[54] D. McAdam and R. Paulsen, “Specifying the relationship between social ties and activism,” *The American Journal of Sociology*, vol. 99, no. 3, p. 640, 1993.

[55] L. Sweeney, “k-anonymity: A model for protecting privacy,” *International Journal of Uncertainty Fuzziness and Knowledge Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[56] P. Golle, “Revisiting the uniqueness of simple demographics in the US population,” in *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society*, 2006, p. 80.

[57] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, p. 3, 2007.

[58] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *Proceedings of the IEEE International Conference on Data Engineering*, April 2007, pp. 106–115.

[59] C. Aggarwal, "On k-anonymity and the curse of dimensionality," in *Proceedings of the 31st International Conference on Very Large Data Bases*, 2005, p. 909.

[60] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *Proceedings of Workshop of Privacy in the Electronic Society*, 2005, pp. 71–80.

[61] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the Facebook," in *Proceedings of the 6th Workshop on Privacy Enhancing Technologies*, 2006, pp. 36–58.

[62] "Police watch your news feed, too," *The Badger Herald*, May 2008. [Online]. Available: http://badgerherald.com/news/2008/05/06/police_watch_your_ne.php

[63] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "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*, February 2010, pp. 251–260.

[64] M. Lucas and N. Borisov, "flybynight: Mitigating the privacy risks of social networking," in *Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society*, 2008, pp. 1–8.

[65] S. Guha, K. Tang, and P. Francis, "NOYB: Privacy in online social networks," in *Proceedings of the 1st Workshop on Online Social Networks*, 2008, pp. 49–54.

[66] W. Luo, Q. Xie, and U. Hengartner, "FaceCloak: An architecture for user privacy on social networking sites," in *Proceedings of the 2009 IEEE International Conference on Privacy, Security, Risk and Trust*, August 2009, pp. 26–33.

[67] A. Felt and D. Evans, "Privacy protection for social networking APIs," *Web 2.0 Security and Privacy*, 2008. [Online]. Available: http://www.cs.berkeley.edu/~afelt/privacybyproxy.pdf

[68] K. Singh, S. Bhola, and W. Lee, "xBook: Redesigning privacy control in social networking platforms," in *Proceedings of the 18th USENIX Security Symposium*, 2009, pp. 249–266.

[69] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proceedings of the 28th IEEE Symposium on Security and Privacy*, 2008, pp. 111–125.

[70] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in *Proceedings of the 30th IEEE Symposium on Security and Privacy*, 2009, pp. 173–187.

[71] J. Bonneau, J. Anderson, R. Anderson, and F. Stajano, "Eight friends are enough: Social graph approximation via public listings," in *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, 2009, pp. 13–18.

[72] L. Backstrom, C. Dwork, and J. Kleinberg, “Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography,” in *Proceedings of the 16th International Conference on World Wide Web*, 2007, p. 190.

[73] E. Jones and H. Gerard, *Foundations of Social Psychology.* New York, NY: John Wiley & Sons Inc, 1967.

[74] D. Bohning, “Multinomial logistic regression algorithm,” *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 1, pp. 197–200, 1992.

[75] K. Jones et al., “A statistical interpretation of term specificity and its application in retrieval,” *Journal of Documentation*, vol. 60, pp. 493–502, 2004.