# Unveiling Privacy Setting Breaches in Online Social Networks

Xin Ruan[1], Chuan Yue[2], and Haining Wang[1]

[1] The College of William and Mary, Williamsburg, VA 23187, USA
{xruan,hnw}@cs.wm.edu
[2] University of Colorado Colorado Springs, Colorado Springs, CO 80918, USA
cyue@uccs.edu

**Abstract.** Users of online social networks (OSNs) share personal information with their peers. To manage the access to one's personal information, each user is enabled to configure its privacy settings. However, even though users are able to customize the privacy of their homepages, their private information could still be compromised by an attacker by exploiting their own and their friends' public profiles. In this paper, we investigate the unintentional privacy disclosure of an OSN user even with the protection of privacy setting. We collect more than 300,000 Facebook users' public information and assess their measurable privacy settings. Given only a user's public information, we propose strategies to uncover the user's private basic profile or connection information, respectively, and then quantify the possible privacy leakage by applying the proposed schemes to the real user data. We observe that although the majority of users configure their basic profiles or friend lists as private, their basic profiles can be inferred with high accuracy, and a significant portion of their friends can also be uncovered via their public information.

## 1 Introduction

Online social network (OSN) websites have attracted a large number of users in the past few years. Facebook, the most popular OSN, was launched in 2004; by March 2013, the monthly active users exceeded 700 million [2]. Each user account typically includes the user's basic profile, such as gender, education, and friend list, and other personal data, such as photos and messages. Clearly not every user is willing to share all its information with peer users, either friends or strangers [18]. Accordingly, many social network sites allow a user to take control over its information visibility by configuring privacy settings. Thus, users are able to set their information visibility to different types, and the setting granularity varies from site to site. For instance, except for profile image and name, a Facebook user is capable of configuring its friend list, each piece of profile information, wall post and photo accessibility to strangers and specific friends.

However, some of an OSN user's private information that is protected by its privacy setting can be easily compromised. In other words, a privacy setting is

T. Zia et al. (Eds.): SecureComm 2013, LNICST 127, pp. 323–341, 2013.

not effective as what it claims to be. This is due to the intrinsic vulnerabilities inside the privacy setting policy. For instance, as shown in Figure 1, user A and user B are mutual friends; each configures its privacy independently such that their information visibility are as the figure shows. An attacker, who does not set up connections with user A or user B, has no access to user A's friend list, but can access some of its photos or posts; thus some of user A's friends, who responded to A's posts or left photo comments, are leaked. When the attacker also visits user B, who has a public friend list, the attacker can confirm the connection between A and B. Exploiting this kind of vulnerability, we wonder whether A's friends or B's basic information could be uncovered even with the protection of their personalized privacy settings. More generally, we attempt to measure, from an average attacker's perspective, with limited resources, how much of a user's privacy could possibly be compromised based on its plainly leaked information.

From the stance of a stranger to a target user, this paper strives to evaluate the user's privacy setting breaches on a large scale and attempts to answer the following questions:

- Can one's privacy setting be undermined by developing more sophisticated and practical schemes, which can infer more private profile information based on what has been directly published from the person's homepage?
- How accurate can users' privacy be inferred? While users can configure their privacy settings to different types, can the amount of inferred privacy be quantified given each privacy setting type?
- Is the amount of inferrable privacy mainly determined by the user's privacy setting? If so, can the number of affected users with a certain setting be estimated on a large scale?

Although previous research [16, 17] has investigated the gap between OSN users' privacy expectation and their actual privacy settings, the vulnerabilities in privacy settings themselves are not studied. Yet there are rare existing research that specifically examines whether a privacy setting can keep the privacy of user information as it is configured. While several efforts [8, 14, 29] have demonstrated the possibility to infer OSN users' one attribute value from another, or to infer the connections, they are based on (1) a large amount of training data [29] or (2) the assumption of the availability of specific kinds of information, such as group membership [14, 29] and music interests [8], which in reality may be set as private by users. The effects of users' privacy settings upon their profiles are not taken into account, let alone to measure the privacy setting breach. A large number of users, who share certain attribute values with the target users, are required as the training data to conduct the information inference. Thus, those strategies can only be taken by attackers with rich resources.

In this paper, we investigate whether certain privacy settings can effectively protect a user's private information as the user configured. We dwell on measuring and quantifying the unintentional leakage of a target user's basic profile information and friend list, which are the pivot of its social profile. For each target user with a certain privacy configuration, we propose the profile and connection inference

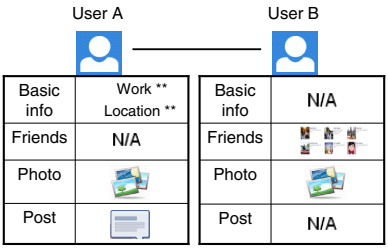


**Fig. 1.** An Attacker's View

schemes based on the user's publicly available information. In addition, instead of relying on a large amount of training data, our approach only needs a small number of users in the target user's neighborhood. The proposed schemes can be conducted by any average users without many resources. We crawl and collect about 300,000 Facebook users' publicly available information as our dataset. The status-quo of those users' privacy settings is measured. Then, we quantify the amount of inferrable private information by using our proposed schemes, and observe that a remarkable amount of privacy could be uncovered, indicating that privacy settings do not effectively guarantee users' information privacy.

The remainder of the paper is organized as follows. Section 2 surveys related work. Section 3 introduces the dataset we collected, and the privacy setting statistics. Section 4 illustrates the privacy breach of each primary setting case under different attack schemes. Section 5 quantifies the breach based on the Facebook dataset. Section 6 discusses the generality of privacy breach in other OSNs, and finally Section 7 concludes the paper.

## 2 Related Work

There are two major research directions on the privacy and security issues in OSNs: (1) to reveal the privacy threats in OSNs by conducting surveys [16, 17] and proposing attack models [26], information inference algorithms [6, 8, 9, 13, 14, 19, 28], de-anonymization algorithms [4, 21], and re-identification algorithms [27]; and (2) to reinforce users' privacy by redesigning the OSN system structure [5, 10, 20, 23] and conducting anonymization [22, 25]. This paper investigates the privacy setting breaches, which belongs to (1). We describe the related work as follows.

The disparity between users' actual privacy settings and their privacy expectation in Facebook has been studied by Madejski et al. [17] and Liu et al. [16]. They obtained users' expectations by conducting surveys and retrieved their factual privacy settings; and then detected the inconsistency between the two. Both found that there was a significant variance between users' privacy expectations and their privacy settings. But they assumed that the privacy setting can effectively protect the data that it is configured to protect. In contrast, this paper intends to challenge this assumption and unveils the privacy setting vulnerability in itself. In addition, we measure the privacy setting status-quo on a much larger scale.

Regarding information inference, there are profile mining [6, 8, 19, 29] and link mining [13–15, 24, 28] approaches, both of which this paper explores. Zheleva et al. [29] presented several classification models using links and group memberships to infer the target users' profiles. But in many OSNs such as Facebook, the group membership is covert by default. Moreover, it assumes that a specific percentage of attribute values are publicly available to perform the inference, and a user set that consists of thousands of users as training data is needed for classification.

Chaabane et al. [8] extracted semantic correlations among users' music interests, and computed each user's probability vector belonging to certain semantic topics. The users with similar vectors shared the same attribute value. However, this method is limited to those users who have published their music interests, and is not applicable to more general users who have not done so. A large dataset is also needed for classification.

Mislove et al. [19] assumed that users sharing the same attribute values were inclined to form dense communities. The traditional community detection algorithm is modified to take user's attribute values into consideration. The algorithm is applied to a school student dataset to infer their majors schools, and etc., but when it is applied to a larger user set from a broader geographical area, the accuracy is much lower than that using the student dataset.

Compared to these related works, this paper designs inference schemes from the stance of an individual user instead of a global view, thus it avoids the need of a large amount of training data and only demands the information of the target user's reachable neighbors. More importantly, our schemes take the actual availability of users' attribute values into consideration, instead of assuming specific attribute values to be in hand.

Another important privacy threat is the compromise of a user's connections, i.e., the friend list. Leroy et al. [14] uncovered the social graph given the user's group membership information. However, it is not easy to obtain these group-related data in most OSNs, in which group information is private. Staddon et al. [24] inferred a user's friend list based on the situation that most OSNs provide the shared friend function once a connection has been set up to the target user. However, the dilemma is if the attacker connects to the target user, likely the target user's friend list is already accessible to the attacker. Bonneau et al. [7] also aimed at uncovering a target user's friend list in Facebook by exploiting the public listing feature, but the feature has been disabled and is not available anymore.

## 3 The Facebook Dataset

Facebook was chosen as our research target because it is the world's most populous OSN providing many flexible features and diverse user resources. More importantly, its privacy setting policy is similar to the policies that most existing OSNs adopt, but in finer granularity. In Facebook, one can set each of its information item individually as "Public," which means to be visible to every user, or visible only to specific or all friends.

While collecting the dataset, the collector acts as a user who neither belongs to any specific group nor sets up connections with any of the sample users.

The retrieved data are all set as "Public," i.e., accessible to every normal user. Hence, the inference experiments can be reproduced by any other users. Moreover, since we only collected public information, none of Facebook's security policies were broken. For privacy concern, user names and IDs are anonymized.

The dataset is organized into a database, consisting of about 300,000 Facebook users. The crawling originated from 50 graduate students at the same institution and was conducted in a breadth-first manner. Out of the total users, about 120,000 users were crawled at the beginning phase, and all their main profile subpages were collected. The rest about 180,000 users were crawled thereafter, and all but their photo subpages were collected as photo pages are not used for evaluation. Out of the 300,000 users, there are 909 users all of whose friends' profiles are also in the dataset; for the rest of users, only some of their friends are in the dataset.

To quantify the information leakage, we emphasize the unintentional revelation of a user's $targetProfile$, including an attribute set: {location, institution} and the friend list. The attribute set is called the basic attribute set, and its element is basic attribute. While $targetProfile$ is the pivot of a user's social profile, other information items from wall like status, messages, to photos are not included in it because they are improvised and hard to infer.

We define the percentage of users that have certain information public as "public ratio." Based on our dataset, the public ratios of users' four main subpages are: 83.8% for profile page, 62.2% for friends page, 55.1% for wall page, and 45.6% for photo page. For a profile page, it is considered to be public when at least one value in the basic attribute set is visible. A photo or wall page is considered to be public if at least one album or post is visible. A friend list is considered public when it is visible.

As many as 37.8% of users conceal their friend lists from strangers. Compared to about 28% for the dataset in Gundecha's work [12], more users in our dataset are aware of connection privacy. Although about 83.8% of users publish one or more basic attribute values, a majority of them provide incomplete basic profiles. Based on the dataset, only 9.9% of users publish complete basic attribute values.

Those statistics demonstrate that a significant number of users customize their $targetProfile$s as private or partially private. The inference of their $targetProfile$s reflects the effectiveness of their privacy settings. Next, we present the schemes to infer each of the two $targetProfile$ items in detail.

## 4 Exploiting Privacy Setting Vulnerability

Targeting a user's $targetProfile$, we design different inference schemes for each possible privacy setting type on the four subpages, including profile, friends, wall, and photo. For easy presentation, the notations we used are listed as follows:

$U$ **: user set.**

$PS(u)$ **:** $u \in U$, user $u$'s privacy setting on four subpages: profile, friends, wall, photo in sequence; denoted as a 4-tuple, and entry value 1 means all basic

**Table 1.** User Sets and ratio

| User Set | $U1$ | $U2$ | $U3$ | | $U4$ | |
|---|---|---|---|---|---|---|
| $PS$ | 0100 | 0001 | 0001 | 1001 | 0000 | 11xx |
| | 0101 | 0010 | 0010 | 1010 | 1000 | |
| | 0110 | 0011 | 0011 | 1011 | | |
| | 0111 | | | | | |
| Ratio | 54.0% | 14.3% | 15.4% | | 22.4% | 8.2% |

attributes are visible in the profile page, visible friend page, some visible posts on the wall or photos, respectively, while 0 represents the opposite.

$BA(u)$ **:** $u \in U$, user $u$'s basic attribute values.

$FL(u)$ **:** $u \in U$, all users in $u$'s friend list, denoted as a user set.

$targetProfile(u)$ : $u \in U$, user $u$'s $targetProfile$, that is $\{BA(u), FL(u)\}$.

$G = (V,\ E)$ : the social graph formed by users in user set $V$, and $E$ consists of the undirectional connections among users in $V$; $\forall u, v \in V$, if $v \in FL(u)$ and $u \in FL(v)$, $(u, v) \in E$. Most frequently it is used to denote a user's neighborhood graph.

$GC(k)$ **:** $1 \leq k \leq n$, a set of members of a community structure detected in a user's neighborhood, and $n$ communities detected in total.

The scenarios under which the $targetProfile$ has to be inferred include when $PS = (0, 1, x, x)$, $PS = (1, 0, x, x)$ and $PS = (0, 0, x, x)$, where $x$ can be either 1 or 0. According to the inference objective and public information, we categorize users into four sets from $U1$ to $U4$ by their $PS$ values. $U1$ and $U2$ consist of users whose $BA$ values can be inferred while $U3$ consists of users whose $FL$ can be inferred from their public information, and $U4$ consists of those whose $BA$ or $FL$ are hard to be directly inferred from their public information.

Table 1 shows the possible $PS$ values in each user set and the ratio of users in it. About 8.2% of users display complete $targetProfile$s to strangers, thus they are not the inference objects. The union of $U1$, $U2$ and $U3$ consists of 69.4% of users, those users' $targetProfile$s are not complete with more or less additional information accessible. In the following subsections, we first illustrate $BA$ inference followed by $FL$; in particular, we infer $BA$ for users in $U1$ and $U2$, then we infer $FL$ for users in $U3$, followed by the hardest case for users in $U4$.

### 4.1 Basic Attributes from Friends

The users in $U1$ display incomplete or no $BA$ but their friend lists are visible, and their $BA$s should be inferred. Table 1 shows that 54% of users belong to $U1$, indicating that a large group of users' privacy are threatened if their $BA$s can be properly compromised. This scenario is formulated as:

$U1 = \{v | v \in U \ and \ PS(v) = (0, 1, x, x)\}$;
Inference objective: $BA(v), v \in U1$;
Public information: $FL(v), v \in U1$.

Intuitively, a user's geographical location, occupation, and interests affect the formation of its social circle. Some connections are set up with colleagues or classmates, while others are from interest communities. Thus, its friends could be classified into different groups, each of which is distinguished by an attribute value shared by the group members and the user. Some of its friends may belong to multiple groups. For example, one author's Facebook friends can be classified into three main groups: one from college, one from graduate school, and one from the current city. Some friends from the graduate school are also in the current city, while no one from college is in the current city. The three groups are distinguished by attribute values at the city or institution level. The friends could be classified into smaller groups by using finer granularity attributes like class and department. The friends in the same group have a higher chance to connect to each other than those from different groups. In other words, community structure exists in the user's friend circle: the connections inside a community are denser than the connections among communities [11].

Therefore, for $v \in U1$, this feature can be exploited to infer $BA(v)$, i.e., to study the connections among $v$'s neighbors and detect communities. We first obtain the social graph in $v$'s neighborhood, $G = (V, E)$ and $V = FL(v)$, by traversing $v$'s friends and retrieving their profile pages and friend lists, although some of them are private. Then, we conduct the community detection in the graph. After that, we identify the most widely shared basic attribute value within each community as the *community feature*, and assemble those features together to form $BA(v)$. During the neighborhood traversal, neither users who have private profiles nor those who have private friend lists are eliminated during the process. This is because their information could be leaked from their shared friends with $v$, who have looser privacy configurations. The steps to infer $BA(v)$ are detailed below as **Scheme 1:**

1. Traverse each user $u$ for $u \in FL(v)$ and retrieve $BA(u)$ and $FL(u)$; then form $v$'s neighborhood graph $G = (V, E)$, $V = FL(v)$, based on $FL(u)$ for each $u \in FL(v)$.
2. Detect the communities in $v$'s neighborhood graph, $G = (V, E)$, $V = FL(v)$, using Girvan-Newman algorithm [11]; and the resulting communities are denoted as $GC(1)$, $GC(2)$, $\cdots, GC(n)$.
3. For each community $GC(k)$, $1 \leq k \leq n$, find the *community feature* $A(k)$ and its frequency such that $A(k) \in BA(u)$ for $u \in GC(k)$ and $A(k)$ is the most widely shared basic attribute value among the community members.
4. Merge $A(k)$s of the same value and sum up their frequencies for $1 \leq k \leq n$; then sort the merged $A(k)$s by institution and location separately in decreasing frequency order. The top-ranked values from the two sorted lists are taken as $BA(v)$.

The Girvan-Newman algorithm is chosen as our community detection algorithm because it does not hold bias against small-sized graphs. Since the detection algorithm is conducted on the $v$'s neighborhood graph, which is on comparatively small scale, the algorithms that hold bias to sparsely connected or small graphs are excluded from our consideration. On the other hand, the

Girvan-Newman algorithm proceeds by removing the edges with the highest edge-betweenness [11] value iteratively, and the procedure is suitable to conduct on small-sized graphs.

As for the number of top values to take in step 4, it can be decided by the target user's number of friends and the frequency of sorted values. More friends indicate more experience, and more values should be taken. Meanwhile, the values whose frequency is comparable with that of the top one value could also be taken. Intuitively, the higher the frequency, the higher the probability the value is accurate.

### 4.2 Basic Attributes from Wall and Photos

The users in $U2$ display incomplete or no $BA$ and conceal their friend lists from strangers, but some of their wall posts or photos are visible. We need to infer their $BA$s. Out of the dataset, 14.3% of the users belong to $U2$. It is formulated as:

$U2 = \{v \mid v \in U \text{ and } PS(v) = (0, 0, x_1, x_2),\ x_1, x_2 = 0, 1 \text{ and } x_1 + x_2 > 0\}$;
**Inference objective** : $BA(v)$, $v \in U2$;
**Public information** : $v$'s public wall posts or photos.

Although the target user $v$'s friend list is private, a direct leakage of $v$'s connections is in $v$'s photos or wall posts where its friends leave comments or get tagged. Different numbers of connections are leaked for different users, depending on their activities and privacy settings on the wall and photo subpages. We randomly choose 330 users in the dataset seeds' neighborhood that belong to $U2$, and crawl their public photos and part of wall posts. The cumulative number of users having less than or equal to a certain number of leaked friends is depicted in Figure 2. While about 90 users have no friends leaked, over half of the users have more than five friends leaked and the maximum number of leaked friends is 295. If all the public wall posts are crawled, the number of leaked friends would increase.

Whereas $v$ has some leaked friends, they may compose a small portion of $v$'s total friends. Namely, the leaked friends can be too spare to form detectable communities in $v$'s neighborhood. Therefore, Scheme 1 is not applicable to users in $U2$. We seek to uncover $BA(v)$ in $v$'s leaked friends' neighborhood, instead of $v$'s neighborhood. First we traverse the directly leaked friends to retrieve their public friend lists and verify their connections to $v$. For those verified friends, their own friends can be traversed to obtain their neighborhood graphs and then detect communities in their neighborhoods. As illustrated before, the *community feature* is supposed to be the most widely shared by community members. Here $v$ is classified to a certain community in each of the verified friends' neighborhood, and it should have a high probability to share the *community feature*. Accordingly, the steps to reveal $BA(v)$ are detailed below as **Scheme 2:**

1. Look through $v$'s wall and photos to retrieve leaked friends.
2. Traverse each leaked friend to retrieve its friend lists if public and verify its connection with $v$.

3. For each verified friend $u$, traverse its friends and detect communities in $u$'s neighborhood using the Girvan-Newman algorithm, resulting in $GC(1)$, $GC(2)$, $\cdots$, $GC(n)$; if $v \in GC(k)$, find the corresponding *community feature* $A(k)$ and its frequency.
4. Merge and sort $A(k)$s, found in $v$'s leaked friends' neighborhoods, in decreasing frequency order and identify $BA(v)$ in the top values.

Intuitively, the more friends leaked, the more *community feature*s can be found to increase the inference accuracy. Figure 2 demonstrates the possibilities of conducting the scheme. However, some users may display their photo and wall subpages but no comments are there; hence no friends are leaked. These cases are treated the same as these users in $U4$.

Besides, Scheme 2 could also be improved by assigning weights to the leaked friends, under the observation that those friends who comment or leave messages to user $v$ might be closer to $v$ than other friends. Higher priority could be given to the *community feature* found in those closer friends.

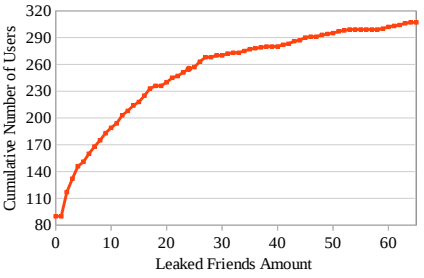


**Fig. 2.** Leaked Friends

**Input:** $R(v)$ = leaked friends
**Output:** $FL(v), C(v)$
**while** $|R(v)| > 0$ **do**
  $R = R(v)$;
  $R(v) = \{\}$;
  **for** $u \in R$ **do**
    Retrieve $FL(u)$;
    $T(v) = T(v) + \{u\}$;
    **if** $FL(u)$ is private **then**
      $C(v) = C(v) + \{u\}$;
    **else**
      **if** $v \in FL(u)$ **then**
        $FL(v) = FL(v) + \{u\}$;
      **for** $w \in FL(u)$ **do**
        **if** $w \in T(v)$ **then**
          pass;
        **else**
          $R(v) = R(v) + \{w\}$;

**Algorithm 1.** Traversal

### 4.3 Friends from Wall and Photos

Those users who conceal friend lists but display some wall posts or photos are categorized into $U3$. We need to infer their $FL$s. As Table 1 shows, 15.4% of users belong to $U3$. The scenario is formulated as:

$U3 = \{v \ | v \in U \text{ and } PS(v) = (x, 0, x_1, x_2),\ x, x_1, x_2 = 0, 1 \text{ and } x_1 + x_2 > 0\}$;
**Inference objective** : $FL(v)$, $v \in U3$;
**Public information** : $v$'s public wall posts or photos.

We aim to uncover $v$'s full friend list while there are some directly leaked friends from $v$'s wall or photo subpages. Therefore, the inference task can be interpreted as traversing near $v$'s neighborhood graph starting from the leaked friends and ascertaining whether those reachable users are $v$'s friends. A few important issues must be considered to make the traversal practical. First, considering that the number of reachable users increases exponentially with the traversal depth,

we should limit the depth so that the traversal is doable. Second, the $v$'s neighborhood graph may be disconnected; thus, if there are components with no starting friends inside, it is arduous to measure the distance between disconnected components in hops by traversing beyond $v$'s neighborhood. We use the word *component* to refer to a connected subgraph within $v$'s neighborhood. Third, for traversed users having private friend lists, it is difficult to distinguish whether they are $v$'s friends.

Taking these practical issues into account, we refrain the traversal from going beyond $v$'s neighborhood graph. The traversal can be conducted in a breadth-first manner, starting from the leaked friends as roots. It proceeds only on those users whose friend lists include $v$, and stops on users whose friend lists exclude $v$. Those traversed users with private friend lists could be gathered together for further verification. Overall, the inference scheme consists of two steps and are detailed below as **Scheme 3:**

1. Traverse the $v$'s neighborhood graph starting from the leaked friends as Algorithm 1 specified.
2. Determine the connectivity between $v$ and traversed users who have private friend lists.

Algorithm 1 uses the following notations:

$R(v)$ **:** the set of users that are yet to be traversed in the coming iteration;
$R$ **:** the set of users that are to be traversed in the current iteration;
$T(v)$ **:** the set of users that have been traversed;
$C(v)$ **:** the set of users that have been traversed but have their friend lists private.

Initially, $R(v)$ consists of the leaked friends from photos and walls, while $T(v)$, $C(v)$, and $FL(v)$ are empty. Each iteration represents the traversal of users a certain depth away from roots. The algorithm terminates when no users traversed in the previous round are friends of $v$, that is $R(v)$ is empty. Furthermore, the algorithm could be adjusted to terminate in advance by confining the traversal depth. The depth can be recorded by counting the number of iterations, and the traversal terminates when the depth limit has been reached.

When the traversal algorithm terminates normally, all of $v$'s friends who have public friend lists and are in the same components with the leaked friends should be included in the derived set $FL(v)$. On the other hand, users who are in different components from the leaked friends cannot be reached. This limitation is due to the feasibility concerns of Scheme 3. However, as the evaluation result in Section 5.2 indicates, on average the largest component in a user's neighborhood consists of over 75% of its friends. In other words, a leaked friend is likely to be included in the largest component; and thus the majority of $v$'s friends are reachable from the leaked friends. Besides, as the component size and edge density vary in $v$'s neighborhood, the traversal complexity differs.

**Complexity of Algorithm 1.** The complexity of algorithm 1 is analyzed in terms of the number of users whose information have to be retrieved. Assume that all users' numbers of friends are at the same magnitude, denoted as $f$. Algorithm 1 constrains the traversal to be within two hops away from

the target user $v$; and thus all $v$'s friends and its friends' friends are traversed in the worst case. We first take the $v$'s $f$ friends into count; and then we count its friends' friends as follows. In the algorithm, each user can only be traversed once. Thus, counting $v$'s friends' friends should exclude $v$'s friends. Let $G = (V, E), V = FL(v)$ denote $v$'s neighborhood graph; and then for each $u \in V$, $f - degree(u)$ of its friends would be counted, which excludes $v$'s friends. Thus, $\sum_{u \in V} f - degree(u)$ more users should be counted, that is, $f^2 - \sum_{u \in V} degree(u)$, in which $\sum_{u \in V} degree(u) = 2|E|$ according to graph theory. In total, the algorithm is in $\mathcal{O}(f + f^2 - 2|E|)$.

Therefore, the more densely $v$'s friends connect to each other, the fewer users have to be traversed. The complexity varies between $\Theta(f^2)$ and $\Theta(f)$. The best case is when $v$'s friends compose a complete graph, i.e. $|E| = \frac{f(f-1)}{2}$, then the complexity is $\mathcal{O}(f)$. When the algorithm terminates by limiting the traversal depth, the complexity would be lower.

As for the second step of Scheme 3, i.e., distinguishing the connectivity between $v$ and traversed users who have private friend lists, the traditional link prediction algorithms such as common friends or Katz [15] can be employed.

### 4.4 No Leaked Friends

The users holding the strictest privacy settings are categorized into $U4$. These users set friends, wall and photo subpages as private and display some or no profile information. The users in this category constitute about 22.4% of the dataset. We need to infer both their $FL$s and $BA$s. While the inference schemes presented before start from some friend connections, the users in $U4$ display none of their friends.

Other means have to be sought to identify possible friends. One source to seek is the special friends or family member sections. Otherwise, the search people function could be exploited by using a user's location or institution, if provided, as keywords. Then, the search results can be traversed one by one to check whether the target user is included in their friend lists. As long as one of the target user's friends with public friend lists can be found, previous schemes can also be conducted to reveal its $targetProfile$. Otherwise, their privacy can not be inferred by our schemes.

In the next section, we apply these schemes to the dataset presented in Section 3 to quantify the privacy that can be compromised in each case.

## 5 Evaluation

The $BA$ inference schemes are conducted on users who display their $BA$ values, and the $FL$ inference schemes are conducted on users who display their $FL$ values; otherwise, the ground truth is not available for verification.

For the $targetProfile$ inference, evaluation bias may be induced in the results when a user's public profile is incomplete or fallacious. Considering the real name policy of Facebook [1], the problem of profile authenticity will not be as significant as incompleteness, which results in false positives. Especially for the

location attribute values, only hometown and current city are available in the ground truth, while schemes 1 and 2 can also infer other cities where a user has ever stayed, such as those associated with the institutions where the user has ever been. Hence, the actual location inference accuracy should be higher than what the results illustrate.

### 5.1 Inferring Basic Attribute Values

Scheme 1 is evaluated first, which can be applied to the users with public friend lists. Out of the dataset,there are 909 users all of whose friends are in the dataset; thus, scheme 1 is applied to those users, referred to as evaluated users. Those who display nothing in their profiles are excluded due to the lack of ground truth for verification. Besides, users with more than 1,000 friends are excluded from the evaluation results. They consist of 5.17% of the total evaluated users, but less than three, if not zero, users fall into each user sample bin in this range; sparsity of user sample isn't likely to result in representative evaluation result.

We use the "igraph" [3] library to detect communities in each evaluated user's neighborhood with the Girvan-Newman algorithm [11]. In each community, the most frequently shared basic attribute value, the *community feature*, can be either a location or an institution value. We identify both the most-shared institution and location values when the community size is above average, and the one with lower frequency is called the *additional feature* of the community. Then we merge and sort those community features and additional features separately in decreasing frequency order by location and institution, respectively. The top ranked values are taken as the user's inferred basic attribute values.

We evaluate the basic attribute inference schemes from the following three aspects. (1) How many basic attribute values could be inferred? The number of public attribute values in evaluated users' homepages which are taken as ground truth, varies from user to user; thus, the number of basic attribute values that can be inferred for each user should be measured. (2) How accurate are inferred values? The number of top values from sorted community features, taken as inferred basic attribute values, can be adjusted; hence the accuracy of each value in the top rank should be be measured. (3) Whether the number of correctly inferred basic attribute values and the inference accuracy are affected by the number of the evaluated user's friends. Since the basic attribute values are inferred from the target user's friends' information, we want to know whether the number of friends affects the inference accuracy or number. Figures 3 to 6 give answers to those questions one by one. In all these figures except for Figure 6, the x-axis value is the number of users' friends and the y-axis value is the average value of users whose number of friends fall into the 20 user sample bin.

Figure 3 depicts the number of correctly inferred basic attribute values compared to the number of basic attribute values in ground truth. The figure shows that more attribute values could be inferred for users with more than 100 friends compared to those with less friends. It verifies the previous claim that the more friends a users has, the more attribute values could be derived; but the differences among users who have more than 120 friends are not significant. On average,

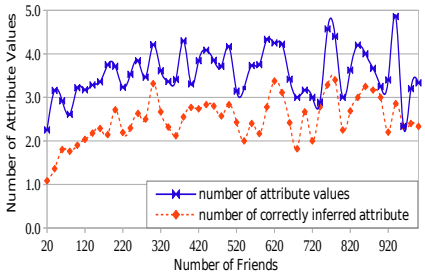


**Fig. 3.** Inferred Attribute Number

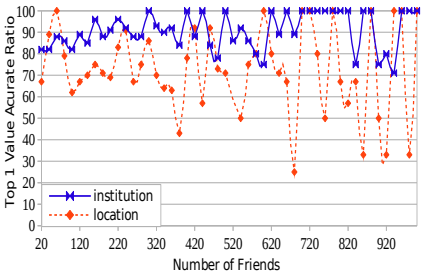


**Fig. 4.** Inference Accuracy

more than two attribute values could be correctly inferred. Attribute values that are not reflected in a user's community features cannot be inferred; one possible reason is that the user is not active in certain OSN communities, or its residence in a certain institution or city is too short to form a community.

The accuracy of the top values taken as inferred basic attribute values are shown in Figures 4 and 5. The accurate ratio is defined as the ratio between the number of verified inferred attribute values and the number of inferred values. In Figure 4 top 1 institution and location are taken as inferred values while in Figure 5 top 2 and top 3 institutions are taken as inferred values.

Figure 4 shows that the inference accurate ratio for institution is about 90% on average, and overall, the more friends the target user has, the higher the average accurate ratio is. Meanwhile the accurate ratio of location is not as good due to the false positives incurred by the incomplete ground truth of location values. As we mentioned at the beginning of this section, only hometown and current city are included in the ground truth for location while we infer all the places that the user has ever been. In addition, the accurate ratio of the top 1 location value for users with more than 500 friends fluctuates more strongly. One reason is that usually the larger the number of friends, the more experience a user has or the more locations a user has ever been, and in turn the less chance for the hometown or current city to be derived as the top 1 inferred location value. Another reason is that users with more than 500 friends are sparse at some point compared to users with fewer friends; thus the accurate ratio cannot be averaged and tends to go extremes due to the sparse user sample. This also explains the higher variance for those users in Figures 3 and 5.

Though the missing of ground truth for location leads to false positives, each institution is usually associated with a location; as long as institutions are correctly inferred, corresponding locations could also be derived. Hence, we further evaluate the accurate ratio of inferred institution information in Figure 5. Figure 5 depicts the accurate ratio of top 2 and top 3 ranked institution values. It shows the accuracy of top 2 institution values is over 80%, which on average is higher than that of top 3 institution values. It verifies our claim that higher-ranked community features hold higher probability to be shared by the target user. Besides, the accurate ratio is not largely affected by the number of users' friends.

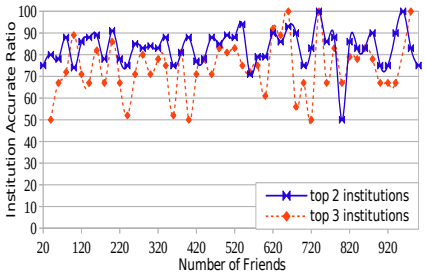


**Fig. 5.** Top Institutions Accuracy

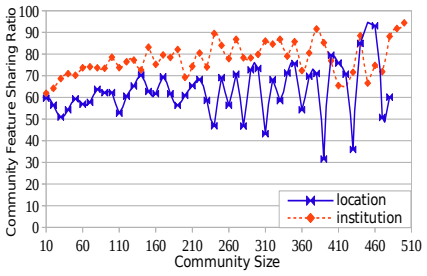


**Fig. 6.** Community Feature Sharing

For users belonging to $U2$, we first measure the community feature sharing ratio to evaluate their basic attribute values inference accuracy, since their basic attributes are derived from the community feature in their leaked friends' neighborhood. Figure 6 depicts the community feature sharing ratio, and x-axis value is the community size. More than 8,500 communities are detected in the evaluated users' neighborhood. On average, the sharing ratio is higher when the community feature is an institution value compared to when it is a location value. This difference can also be explained by the ground truth incompleteness of location information. Though the community features are not 100% shared by all members, they will not be directly taken as the inferred basic attribute values and the wrong community features will be eliminated in the later steps of Scheme 2.

We further evaluate the inference accuracy of Scheme 2 on some of the dataset's seed users which belong to $U2$. Because seed users are from the same institution and location, the ground truth scraped from users' homepages are complemented by that fact. We detect those seed users' community memberships in their friends' neighborhood, and take the top ranked community features as their inferred attribute values. As a result, the inference accuracy of top 1 ranked feature is 100%.

In summary, for users who conceal their basic attribute values but have their friend list public or some friends leaked from other profile sections, those value could be uncovered with high accuracy by exploiting their friends' information.

### 5.2 Inferring Friend List

For a user $v$ in $U3$, $v$'s retrievable friends, according to Scheme 3, are confined to those who are in the same component with one of the leaked friends. As defined in Section 4.3, a component is a connected subgraph within $v$'s neighborhood. We first measure the components in users' neighborhoods. Based on the evaluated users, most of their neighborhood graphs are disconnected, on average 20 components exist and the number of components increases with the number of a user's friends. While there are a noticeable number of components, most of them are small. Figure 7 illustrates the ratio of a user's friends that are in their largest neighborhood component, over 85% of friends on average are included in

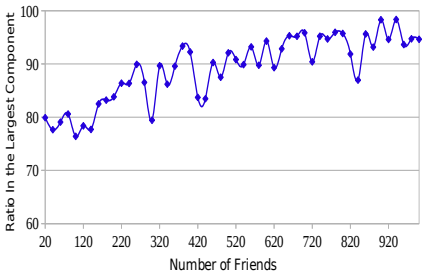


**Fig. 7.** Friends in the Largest Component

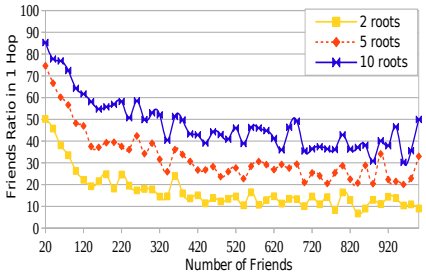


**Fig. 8.** Traversed Friends Ratio in 1 Hop

the largest component. The more friends a user has, the larger portion of friends are in the largest component. Thus, as the leaked friends are likely to be in the largest component, a majority of friends could be reached from them.

In Figure 8, the ratio of traversed friends in the evaluated users' neighborhoods is illustrated, and the traversal starts from different number of roots in one hop away. Each curve represents a different number of roots, which are randomly chosen from target user's friends. For users with fewer than 100 friends, a majority of friends could be traversed in one hop from five roots, while for users with more friends, about 10%, 25%, and 35% of friends could be traversed in one hop away from two, five, and ten roots, respectively. Over all, the more friends a user has, the more of its friends can be reached via traversal given the same number of roots and hops.

Figure 9 indicates the ratio of friends traversed in two hops away. About 70% of friends could be traversed from 5 roots, and 80% of friends could be traversed from 10 roots. The curve for two roots fluctuates more violently because the choice of roots affects the traversal path and a high-degree node results in more retrieved friends. When starting from 5 or 10 roots, the high-degree nodes stand a higher chance to be traversed as roots or within two hops. Still, on average about half of a user's friends could be retrieved from two randomly chosen roots in two hops. Interestingly, the ratio is not clearly affected by users' number of friends. It means that no matter how many friends a user has, most of its friends are closely connected while some are estranged from others.

To sum up, for users who conceal their friend lists but display other profile sections from which some of their friends could be leaked out, over half of their friends could be revealed using our traversal algorithm starting from the leaked friends in two hops. The complexity of the traversal algorithm ensures the traversal can be conducted in limited resource.

After that, we measure the second step of scheme 3, i.e., to distinguish the connections between user $v$ and the traversed users who have private friend lists. Those users are those who connected to $v$'s friends and have private friend lists. The number of common friends is taken as the metric to infer the connections. Those private-friend-listed users are sorted by their numbers of friends shared with $v$, which is leaked from $v$'s public-friend-listed friends. The top quarter of users are taken as $v$'s hidden friends. Figure 10 illustrates the inference accuracy,

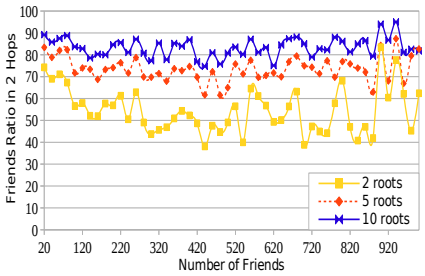


**Fig. 9.** Traversed Friends Ratio in 2 Hops

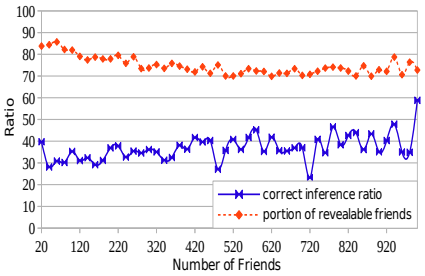


**Fig. 10.** Private-Friends Inference Ratio

and it also illustrates the total revealable friends ratio, which consists of both the public-listed friends and those hidden friends. Compared to the results of [15] which also used common neighbors as the metric to infer co-authorship, our accuracy is slightly higher. In total, for users belonging to $U3$, more than 70% of their friends could be correctly revealed on average by Scheme 3.

Users in $U4$ hide all connections, which is hardest to infer their $targetProfile$. However, if some of their friends are known beforehand or can be found by using the search people function mentioned in Section 4.4, their $targetProfile$ can be inferred and evaluated similar as stated above.

## 6 Discussion

While our approach explores a user's information visibility from the perspective of a stranger, it cannot know the privacy customization to the user's friends. However, the privacy setting for strangers can only be stricter than that for friends. In other words, friends must be able to access more information than strangers. Thus, if some private information could be correctly inferred by a stranger, the inference can also be reproduced by friends.

If a user does not post certain profile item on Facebook such as education, we cannot know whether the invisibility is due to privacy setting or vacancy. However, if the inferred information could be verified based on the ground truth retrieved from other sources, we still view such a case as privacy leakage.

Due to the lack of ground truth, the experiments are only conducted on users who display their $targetProfile$s to strangers. However, we speculate that those users with stricter privacy are also inclined to be more prudent in setting up connections. Thus, their online friend circles are created in a more moderate manner, which does not increase the difficulty of community feature detection or neighborhood graph traversal. Therefore, our evaluation results reflect a possible privacy breach of average users.

The profile inference schemes proposed in this paper are not limited to Facebook. They could also be applied to other OSNs that enable privacy configuration and allow users to post a variety of data other than profile and connection. Those OSNs include MySpace, Google+, and Renren, in which users could also upload photos,

leave messages or comments, and customize the visibility of different types of information. When the accessibility of a user's profile or connections is constrained, the information revelation could be initiated from public connections in the friend list or posts from friends by using our schemes 1, 2 or 3.

## 7 Conclusion

In this paper, we investigated the unintentional privacy disclosure of OSN users even with the protection of privacy settings. We first examined users' privacy settings on different information sections of a large dataset collected from Facebook. Then, for each possible privacy configuration, we proposed corresponding schemes to reveal basic profile and connection information starting from leaked public connections on the target user's OSN homepage. Finally, using our dataset, we quantified the achievable privacy exposure in each case, and measured the accuracy of our privacy inference schemes given a different amount of public information. The evaluation results indicate that a user's private basic profile could be inferred with high accuracy, while a user's covert connections could be uncovered in a significant portion based on even a small number of directly leaked connections.

Our privacy inference schemes can be conducted by attackers without much resources; and those schemes are applicable to users adopting specific privacy settings. The dataset statistics show that a majority of users are among that group. Therefore, the privacy of those users could be undermined facilely and the actual information privacy level of them may fail to meet what their privacy configuration specifies. We discussed that our privacy inference schemes could be applied to other OSNs that provide similar features as Facebook. We plan to analyze the privacy breach on those OSNs in the future.

**Acknowledgement.** We would like to thank the anonymous reviewers for their insightful feedback. This work was partially supported by ARO grant W911NF-11-1-0149.

## References

[1] Facebook name policy, `http://www.facebook.com/help/?page=258984010787183`
[2] Facebook newsroom, `http://newsroom.fb.com/`
[3] IGRAPH, `http://igraph.sourceforge.net/`
[4] Backstrom, L., Dwork, C., Kleinberg, J.: Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In: Proceedings of the 16th WWW 2007 (2007)
[5] Baden, R., Bender, A., Spring, N., Bhattacharjee, B., Starin, D.: Persona: an online social network with user-defined privacy. In: Proceedings of the 2009 ACM SIGCOMM (2009)
[6] Balduzzi, M., Platzer, C., Holz, T., Kirda, E., Balzarotti, D., Kruegel, C.: Abusing social networks for automated user profiling. In: Jha, S., Sommer, R., Kreibich, C. (eds.) RAID 2010. LNCS, vol. 6307, pp. 422–441. Springer, Heidelberg (2010)

[7] Bonneau, J., Anderson, J., Anderson, R., Stajano, F.: Eight friends are enough: social graph approximation via public listings. In: Proceedings of the 2nd ACM EuroSys Workshop on SNS 2009 (2009)
[8] Chaabane, A., Acs, G., Kaafar, M.A.: You are what you like! information leakage through users' interests. In: Proceedings of the 19th NDSS 2012 (2012)
[9] Eyal, R., Kraus, S., Rosenfeld, A.: Identifying missing node information in social networks. Artificial Intelligence, 1166–1172 (2011)
[10] Feldman, A.J., Blankstein, A., Freedman, M.J., Felten, E.W.: Social networking with frientegrity: Privacy and integrity with an untrusted provider. In: The 21st USENIX Security 2012 (August 2012)
[11] Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99(12), 7821–7826 (2002)
[12] Gundecha, P., Barbier, G., Liu, H.: Exploiting vulnerability to secure user privacy on a social networking site. In: Proceedings of the 17th ACM KDD 2011 (2011)
[13] Korolova, A., Motwani, R., Nabar, S.U., Xu, Y.: Link privacy in social networks. In: Proceedings of the 17th ACM CIKM 2008 (2008)
[14] Leroy, V., Cambazoglu, B.B., Bonchi, F.: Cold start link prediction. In: Proceedings of the 16th ACM KDD 2010 (2010)
[15] Liben-Nowell, D., Kleinberg, J.: The link prediction problem for social networks. In: Proceedings of the 12th CIKM 2003 (2003)
[16] Liu, Y., Gummadi, K.P., Krishnamurthy, B., Mislove, A.: Analyzing facebook privacy settings: user expectations vs. reality. In: Proceedings of the 2011 ACM SIGCOMM IMC 2011 (2011)
[17] Madejski, M., Johnson, M., Bellovin, S.M.: A study of privacy setting errors in an online social network. In: Proceedings of SESOC 2012 (2012)
[18] Mashima, D., Sarkar, P., Shi, E., Li, C., Chow, R., Song, D.: Privacy settings from contextual attributes: A case study using google buzz. In: PerCom Workshops, pp. 257–262. IEEE (2011)
[19] Mislove, A., Viswanath, B., Gummadi, K.P., Druschel, P.: You are who you know: inferring user profiles in online social networks. In: Proceedings of the 3rd ACM WSDM 2010 (2010)
[20] Mondal, M., Viswanath, B., Clement, A., Druschel, P., Gummadi, K.P., Mislove, A., Post, A.: Limiting large-scale crawls of social networking sites. SIGCOMM Computer Communication Review 41(4), 398–399 (2011)
[21] Narayanan, A., Shmatikov, V.: De-anonymizing social networks. In: Proceedings of 30th IEEE Symposium on Security and Privacy, S&P 2009 (May 2009)
[22] Pedarsani, P., Grossglauser, M.: On the privacy of anonymized networks. In: Proceedings of the 17th ACM KDD 2011 (2011)
[23] Singh, K., Bhola, S., Lee, W.: xbook: redesigning privacy control in social networking platforms. In: Proceedings of the 18th USENIX Security Symposium, SSYM 2009. USENIX Association, Berkeley (2009)
[24] Staddon, J.: Finding "hidden" connections on linkedin an argument for more pragmatic social network privacy. In: Proceedings of the 2nd ACM Workshop AISec 2009 (2009)
[25] Tai, C.-H., Yu, P.S., Yang, D.-N., Chen, M.-S.: Privacy-preserving social network publication against friendship attacks. In: Proceedings of the 17th ACM KDD 2011 (2011)

[26] Wondracek, G., Holz, T., Kirda, E., Kruegel, C.: A practical attack to de-anonymize social network users. In: Proceedings of the 2010 IEEE Symposium on Security and Privacy, S&P 2010 (2010)
[27] Yang, Y., Lutes, J., Li, F., Luo, B., Liu, P.: Stalking online: on user privacy in social networks. In: Proceedings of the Second ACM CODASPY 2012, New York, NY, USA (2012)
[28] Ying, X., Wu, X.: On link privacy in randomizing social networks. In: Theeramunkong, T., Kijsirikul, B., Cercone, N., Ho, T.-B. (eds.) PAKDD 2009. LNCS, vol. 5476, pp. 28–39. Springer, Heidelberg (2009)
[29] Zheleva, E., Getoor, L.: To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In: Proceedings of the 18th WWW 2009 (2009)