A Framework for Computing the Privacy Scores of Users in Online Social Networks

Kun Liu
Yahoo! Labs
Santa Clara, CA
kun@yahoo-inc.com

Evimaria Terzi
Department of Computer Science
Boston University, Boston, MA
evimaria@cs.bu.edu

***Abstract*—A large body of work has been devoted to address corporate-scale privacy concerns related to social networks. The main focus was on how to share social networks owned by organizations without revealing the identities or sensitive relationships of the users involved. Not much attention has been given to the privacy risk of users posed by their information-sharing activities.**

In this paper, we approach the privacy concerns arising in online social networks from the individual users' viewpoint: we propose a framework to compute a privacy score of a user, which indicates the potential privacy risk caused by his participation in the network. Our definition of privacy score satisfies the following intuitive properties: the more *sensitive* the information revealed by a user, the higher his privacy risk. Also, the more *visible* the disclosed information becomes in the network, the higher the privacy risk. We develop mathematical models to estimate both *sensitivity* and *visibility* of the information. We apply our methods to synthetic and real-world data and demonstrate their efficacy and practical utility.

***Keywords*-privacy score; social network; item response theory**

I. Introduction

Social-networking sites have grown tremendously in popularity in recent years. Services such as Facebook and MySpace allow millions of users to create online profiles and share details of their personal lives with vast networks of friends, and often, strangers. As the number of users of these sites and the number of sites themselves explode, securing individuals' privacy to avoid threats such as *identity theft* and *digital stalking* becomes an increasingly important issue.

Unfortunately, even sophisticated users who value privacy will often compromise it to improve their digital presence in the virtual world. They know that loss of control over their personal information poses a long-term threat, but they cannot assess the overall and long-term risk accurately enough to compare it to the short-term gain. Even worse, setting the privacy controls in online services is often a complicated and time-consuming task that many users feel confused about and usually skip.

Past research on privacy and social networks (*e.g.*, [1], [3], [4], [7], [9]) mainly focuses on corporate-scale privacy concerns, *i.e.*, how to share a social network owned by an organization without revealing the identity or sensitive relationships among the registered users. Not much attention has been given to individual users' privacy risk posed by their information-sharing activities.

In this paper, we address the privacy issue from the user's perspective: we propose a framework that estimates a *privacy score* for each user. This score measures the user's potential privacy risk due to his online information-sharing behaviors. With this score, a user can monitor his privacy risk in real time and compare it with the rest of the population to see where he stands. In the case where the overall privacy risks of a user's social graph are lower than that of the user, the system can recommend the user stronger privacy settings based on the information from his social neighbors. Our ultimate objective is to enhance public awareness of privacy, and to help users to easily manage their information sharing in social networks.

From the technical point of view, our definition of *privacy score* satisfies the following intuitive properties: The score increases with the i) *sensitivity* of the information being revealed and ii) with the *visibility* of the revealed information gets in the network. We develop mathematical models to estimate both *sensitivity* and *visibility* of the information, and we show how to combine these two factors in the calculation of the *privacy score*.

Contribution: To the best of our knowledge, we are the first to provide an intuitive and mathematically sound methodology for computing users' *privacy scores* in online social networks. The two principles stated above are rather general, and many models would be able to satisfy them. In addition, the specific model we proposed in this paper exhibits two extra advantages: i) it is container independent, meaning that scores calculated for users belonging to different social networks (*e.g.*, Facebook, LinkedIn and MySpace) are comparable, and ii) it fits the real data. Finally, we give algorithms for the computation of privacy score that scale well and indicative experimental evidence of the efficacy of our framework.

Overview of the technical framework: For a social-network user we compute his *privacy score* as a combination of the partial privacy scores of each one of his profile items, e.g., *user's real name, email, hometown, mobile-phone number, relationship status, sexual orientation, IM*

*This work was done when both authors were at IBM Almaden Research.

screen name, etc. The contribution of each profile item in the total score depends on the *sensitivity* of the item and the *visibility* it gets due to the user's privacy settings. Therefore, the primary input to our framework is an $n \times N$ *response* matrix that shows the privacy settings of each one of the N users for each one of the n profile items. The values appearing in the response matrix are natural numbers; the higher the value of the cell $\mathbf{R}(i, j)$, the more willing user j is to disclose information about item i. Our approach explores this information to compute the privacy scores of users. We do so by employing theories from Item Response Theory (IRT) [2].

In this paper, we do not consider how to conduct inference attacks to derive hidden information about a user based on his publicly disclosed data. We deem this inference problem as an important, albeit orthogonal to our work. Some profile items such as *hobbies* are composite since they may contain many different sensitive information. We decompose this kind of items into primitive ones. Again determining the granularity of the profile items is considered an orthogonal issue to the problem we study here.

Organization of the material: After the presentation of the related work in Section II and the description of the notational conventions in Section III, we present our definitions of privacy score in Section IV. Models and algorithmic solutions are described in Sections V, VI and VII. Experimental comparison of our methods is given in Section VIII. We conclude in Section IX.

II. Related work

To the best of our knowledge, we are the first to present a framework that formally quantifies the privacy risk of online social-network users. What we consider as the most relevant work is the work on scoring systems for measuring *popularity*, *creditworthiness*, *trustworthiness*, and *identity verification*. We briefly describe these scores here.

QDOS score: Garlik, a UK-based company, launched a system called QDOS [1] for measuring people's digital popularity. The primary purpose is to encourage people to share more information to enhance their rankings, which is opposite to ours. Their algorithms are not disclosed.

Credit score: A credit score [2] is used to estimate the likelihood that a person will default on a loan. The credit score is different from our privacy-risk score not only because it serves different purposes but also because the input data used for estimating the two scores as well as the estimation methods themselves are different.

Trust score: A trust score is a measure of how a member of a group is trusted by the others (*e.g.*, [6]). Trust scores could be used by social-network users to determine who can view their personal information. However, our system is used to quantify the privacy risk after the information has been shared.

Identity score: An identity score [3] is used by financial-service firms for tagging and verifying the legitimacy of one's public identity. Our privacy risk score is different from identity score since it serves a different purpose.

III. Preliminaries

We assume there exists a social-network $\mathcal{G}$ that consists of N nodes, every node $j \in \{1, \ldots, N\}$ being associated with a user of the network. Every user has a profile consisting of n profile items. For each profile item, users set a *privacy level* that determines their willingness to disclose information associated with this item. The $n \times N$ *response matrix* $\mathbf{R}$ stores the privacy levels of all N users for all n profile items; $\mathbf{R}(i, j)$ refers to the privacy setting of user j for item i. If the entries in $\mathbf{R}$ take values in $\{0, 1\}$, we say that $\mathbf{R}$ is a *dichotomous*. Otherwise, if the entries in $\mathbf{R}$ take any non-negative integer values in $\{0, 1, \ldots, \ell\}$ we say that matrix $\mathbf{R}$ is *polytomous*.

In a dichotomous response matrix $\mathbf{R}$, $\mathbf{R}(i, j) = 0$ means that user j has made the information associated with profile item i private, whereas $\mathbf{R}(i, j) = 1$ means that j has made item i publicly available. In a polytomous response matrix, $\mathbf{R}(i, j) = 0$ means that user j keeps profile item i private; whereas $\mathbf{R}(i, j) = k$ with $k \geq 1$ means that j discloses information regarding item i to users that are at most k-links away in $\mathcal{G}$.

In general, $\mathbf{R}(i, j) \geq \mathbf{R}(i', j)$ means that j has more conservative privacy settings for item i' than item i. The i-th row of $\mathbf{R}$, denoted by $\mathbf{R}_i$, represents the settings of all users for profile item i. Similarly, the j-th column of $\mathbf{R}$, denoted by $\mathbf{R}^j$, represents the profile settings of user j.

We often consider users' settings for different profile items as random variables described by a probability distribution. In such cases, the observed response matrix $\mathbf{R}$ is just a sample of responses that follow this probability distribution. For dichotomous response matrices, we use P_{ij} to denote the probability that user j selects $\mathbf{R}(i, j) = 1$. That is, $P_{ij} = \text{Prob}\big\{\mathbf{R}(i, j) = 1\big\}$. In the polytomous case, we use P_{ijk} to denote the probability that user j sets $\mathbf{R}(i, j) = k$. That is, $P_{ijk} = \text{Prob}\big\{\mathbf{R}(i, j) = k\big\}$.

In order to allow the readers to build intuition, we start by defining the privacy score for dichotomous response matrices in Section IV, and present algorithms for computing it in Sections VII and V. Polytomous settings are handled in Section VI.

IV. Privacy score in dichotomous settings

The *privacy score* of a user is an indicator of his potential privacy risk; the higher the score of a user, the higher the threat to his privacy. Our basic premises for the calculation

[1] http://www.qdos.com

[2] http://www.myfico.com

[3] http://en.wikipedia.org/wiki/Identity_score

of privacy score are the following: (a) the more sensitive the information revealed by a user, the higher his privacy score; and (b) the wider the information about a user spreads, the higher his privacy score. Therefore, the privacy score of a user is a *monotonically increasing* function of two parameters: the *sensitivity* of the profile items and the *visibility* these items get.

Sensitivity of a profile item: The sensitivity of item $i \in \{1, \ldots, n\}$ is denoted by β_i. This property depends on the item itself. Some profile items are, by nature, more sensitive than others.

Visibility of a profile item: The visibility of a profile item i that belongs to user j is denoted by $\mathrm{V}(i,j)$. It captures how known this item becomes in the network; the widely it spreads, the higher the visibility. Naturally, $\mathrm{V}(i,j)$ depends on the value $\mathbf{R}(i,j)$, which is the explicit privacy level chosen by the user. Thus, in the dichotomous case, we can simply define $\mathrm{V}(i,j) = \mathbf{I}_{(\mathbf{R}(i,j)=1)}$, where $\mathbf{I}_{\text{condition}}$ is an indicator variable that becomes 1 when "condition" is true. We call this the *observed visibility* for item i that belongs to user j. From a statistics point of view, one can assume that $\mathbf{R}$ is a sample from a probability distribution over all possible response matrices. Then, the *true visibility* is computed as $\mathrm{V}(i,j) = P_{ij} \times 1 + (1 - P_{ij}) \times 0 = P_{ij}$, where $P_{ij} = \text{Prob}\big\{\mathbf{R}(i,j) = 1\big\}$. Probability P_{ij} depends both on the item i and the user j.

Privacy score of a user: The privacy score of individual j due to item i, denoted by $\textsc{Pr}(i,j)$, is defined as $\textsc{Pr}(i,j) = \beta_i \bigotimes \mathrm{V}(i,j)$. Operator $\bigotimes$ is used to represent any arbitrary combination function that respects the fact that $\textsc{Pr}(i,j)$ is monotonically increasing with both sensitivity and visibility. For simplicity, we use the product operator.

In order to evaluate the overall privacy score of user j, denoted by $\textsc{Pr}(j)$, we can sum the privacy score of j due to different items.[4] That is,

$$\textsc{Pr}(j) \quad = \quad \sum_{i=1}^{n} \textsc{Pr}(i,j) = \sum_{i=1}^{n} \beta_i \times \mathrm{V}(i,j)\,. \qquad (1)$$

In the above, the privacy score can be computed using either *observed visibility* or *true visibility*. For the rest of the discussion we use the true visibility since we believe that the specific privacy settings of a user are just an instance of his possible setting described by probability distribution $\textsc{Pr}(i,j)$.

In the next two sections we show how to compute the privacy score of a user in a social network based on his privacy settings on his profile items. .

V. IRT-based Computation of Privacy Score: Dichotomous Case

We first introduce some basic concepts from Item Response Theory (IRT) [2]. Then, we show how these concepts are applicable in our setting.

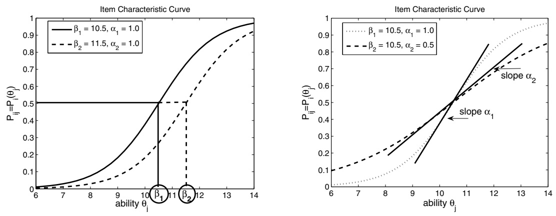
A. Introduction to IRT

IRT has its origins in psychometrics where it is used to analyze data from questionnaires and tests. The goal there is to measure the abilities of the examinees, the difficulty of the questions and the probability of an examinee to correctly answer a given question.

Every examinee j is characterized by his ability level θ_j, $\theta_j \in (-\infty, \infty)$. Every question q_i is characterized by a pair of parameters $\xi_i = (\alpha_i, \beta_i)$. Parameter β_i, $\beta_i \in (-\infty, \infty)$, represents the *difficulty* of q_i. Parameter α_i, $\alpha_i \in (-\infty, \infty)$, quantifies the *discrimination power* of q_i. The intuitive meaning of these two parameters will become clear shortly. The basic random variable of the model is the response of examinee j to a particular question q_i. If this response is marked as either "correct" or "wrong" (dichotomous response), then the probability that j answers q_i **correctly** is given by

$$P_{ij} \quad = \quad \frac{1}{1 + e^{-\alpha_i(\theta_j - \beta_i)}}. \qquad (2)$$

Thus, P_{ij} is a function of parameters θ_j and $\xi_i = (\alpha_i, \beta_i)$. For a given question q_i with parameters $\xi_i = (\alpha_i, \beta_i)$, the plot of the above equation as a function of θ_j [5] is called the *Item Characteristic Curve* (ICC).



(a) $\beta_1 < \beta_2$.

(b) $\alpha_1 > \alpha_2$.

Figure 1. Item Characteristic Curves (ICC); y-axis: $P_{ij} = P_i(\theta_j)$ for different β values (Figure 1(a)) and α values (Figure 1(b)). x-axis: ability level θ_j.

The ICCs obtained for different values of parameters $\xi_i = (\alpha_i, \beta_i)$ are given in Figures 1(a) and 1(b). These figures make the intuitive meaning of parameters α_i and β_i easier to explain.

Figure 1(a) shows the ICCs obtained for two questions q_1 and q_2 with parameters $\xi_1 = (\alpha_1, \beta_1)$ and $\xi_2 = (\alpha_2, \beta_2)$ such that $\alpha_1 = \alpha_2$ and $\beta_1 < \beta_2$. Parameter β_i, the item difficulty, is defined as the point on the ability scale at which $P_{ij} = 0.5$. We can observe that IRT places β_i and θ_j on the same scale (see the x-axis of Figure 1(a)) so that they can compare. If an examinee's ability is higher than the difficulty of the question, then he has better chance to get the answer

[4] Again, any combination function can be employed to combine the per-item privacy scores. For simplicity, we use summation operator here.

[5] We can represent P_{ij} by $P_i(\theta_j)$ to indicate the dependency on θ_j. But in general, we use P_{ij} and $P_i(\theta_j)$ interchangeably.

right, and vice versa. This also indicates a very important feature of IRT called *group invariance*, that is, the item's difficulty is a property of the item itself, not of the people that responded to the item. We will elaborate on this in the experiments section.

Figure 1(b) shows the ICCs obtained for two questions q_1 and q_2 with parameters $\xi_1 = (\alpha_1, \beta_1)$ and $\xi_2 = (\alpha_2, \beta_2)$ such that $\alpha_1 > \alpha_2$ and $\beta_1 = \beta_2$. Parameter α_i, the item discrimination, is proportional to the slope of $P_{ij} = P_i(\theta_j)$ at the point where $P_{ij} = 0.5$; the steeper the slope, the higher the discriminatory power of a question, meaning that this question can well differentiate among examinees whose abilities are below and above the difficulty of this question.

In our IRT-based computation of the privacy score, we estimate the probability Prob$\{\mathbf{R}(i,j) = 1\}$ using Equation (2). However, we do not have examinees and questions, but we rather have users and profile items. Thus, each examinee is mapped to a user, and each question is mapped to a profile item. The ability of an examinee corresponds to the *attitude* of a user: for user j, his attitude θ_j quantifies how concerned j is about his privacy; low values of θ_j indicate a conservative/introvert user, while high values of θ_j indicate a careless/extrovert user. We use the difficulty parameter β_i to quantify the *sensitivity* of profile item i. In general, parameter β_i can take any value in $(-\infty, \infty)$. In order to maintain the monotonicity of the privacy score with respect to items' sensitivity we need to guarantee that $\beta_i \geq 0$ for all $i \in \{1, \ldots, n\}$. This can be easily handled by shifting all items' sensitivity values by a big constant value.

In the above mapping, parameter α_i is ignored. Due to space constraints, we cannot elaborate on this topic, but the general idea is that this parameter allows us to do a finer-grained analysis of items and users.

For computing the privacy score we need to compute the sensitivity β_i for all items $i \in \{1, \ldots, n\}$ and the probabilities $P_{ij} = $ Prob$\{\mathbf{R}(i,j) = 1\}$, using Equation (2). For the latter computation, we need to know all the parameters $\xi_i = (\alpha_i, \beta_i)$ for $1 \leq i \leq n$ and θ_j for $1 \leq j \leq N$. In the following sections, we show how we can estimate these parameters using as input the response matrix $\mathbf{R}$ and employing *Maximum Likelihood Estimation* (MLE) techniques. All these techniques exploit the following three independence assumptions that are inherent in IRT models: (i) independence between items; (ii) independence between users; and (iii) independence between users and items. Experiments in Section VIII show that parameters learned based on these assumptions fit the real-world data very well. We refer to the privacy score computed using these methods as the **Pr_IRT** score.

B. IRT-based computation of sensitivity

In this section, we show how to compute the sensitivity β_i of a particular item i. [6] Since items are independent, the computation of parameters $\xi_i = (\alpha_i, \beta_i)$ is done separately for every item; thus all methods are highly parallelizable.

In Section V-B1 we first show how to compute ξ_i assuming that the attitudes of the N individuals $\vec{\theta} = (\theta_1, \ldots, \theta_N)$ are given as part of the input. The algorithm for the computation of items' parameters when attitudes are not known is discussed in Section V-B2.

1) Item-parameters estimation: The maximum likelihood estimation of $\xi_i = (\alpha_i, \beta_i)$ sets as our goal to find ξ_i such that the *likelihood function*

$$\prod_{j=1}^{N} P_{ij}^{\mathbf{R}(i,j)} \left(1 - P_{ij}\right)^{1-\mathbf{R}(i,j)} \tag{3}$$

is maximized. Recall that P_{ij} is evaluated as in Equation (2) and depends on α_i, β_i and θ_j.

The above likelihood function assumes different attitude per user. In practice, online social-network users form a grouping that partitions the set of users $\{1, \ldots, N\}$ into K non-overlapping groups $\{F_1, \ldots, F_K\}$ such that $\bigcup_{g=1}^{K} F_g = \{1, \ldots, N\}$. Let θ_g be the attitude of group F_g (all members of F_g share the same attitude θ_g) and $f_g = |F_g|$. Also, for each item i let r_{ig} be the number of people in F_g that set $\mathbf{R}(i,j) = 1$, that is, $r_{ig} = \left|\left\{j \mid j \in F_g \text{ and } \mathbf{R}(i,j) = 1\right\}\right|$. Given such grouping, the likelihood function can be written as $\prod_{g=1}^{K} \binom{f_g}{r_{ig}} \left[P_i(\theta_g)\right]^{r_{ig}} \left[1 - P_i(\theta_g)\right]^{f_g - r_{ig}}$. After ignoring the constants, the corresponding log-likelihood function is

$$L = \sum_{g=1}^{K} \left[r_{ig} \log P_i(\theta_g) + (f_g - r_{ig}) \log\left(1 - P_i(\theta_g)\right)\right]. \tag{4}$$

Our goal is now to find item parameters $\xi_i = (\alpha_i, \beta_i)$ to maximize this log-likelihood function. For this we use the Newton-Raphson method [8]. The Newton-Raphson method is a numerical algorithm that, given partial derivatives $L_1 = \frac{\partial L}{\partial \alpha_i}$, $L_2 = \frac{\partial L}{\partial \beta_i}$, $L_{11} = \frac{\partial^2 L}{\partial \alpha_i^2}$, $L_{22} = \frac{\partial^2 L}{\partial \beta_i^2}$, and $L_{12} = L_{21} = \frac{\partial^2 L}{\partial \alpha_i \beta_i}$, it estimates parameters $\xi_i = (\alpha_i, \beta_i)$ iteratively. At iteration $(t+1)$, the estimates of the parameters α_i, β_i denoted by $\begin{bmatrix} \widehat{\alpha_i} \\ \widehat{\beta_i} \end{bmatrix}_{t+1}$, are computed from the corresponding estimates at iteration t as follows:

$$\begin{bmatrix} \widehat{\alpha_i} \\ \widehat{\beta_i} \end{bmatrix}_{t+1} = \begin{bmatrix} \widehat{\alpha_i} \\ \widehat{\beta_i} \end{bmatrix}_{t} - \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}_{t}^{-1} \times \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}_{t}. \tag{5}$$

Discussion: The overall process starts with the partitioning of the set of N users into K groups based on users' attitude. This routine implements an 1-dimensional cluster-

[6] The value of α_i, for the same item, is obtained as a byproduct of this computation.

ing, and can be done optimally using dynamic programming in $\mathcal{O}\left(N^2K\right)$ time. The result of this procedure is a grouping of users into K groups $\{F_1, \ldots, F_K\}$, with group attitudes θ_g, $1 \leq g \leq K$. Given this grouping, the values of f_g and r_{ig} for $1 \leq i \leq n$ and $1 \leq g \leq K$ are computed. These computations take time $\mathcal{O}\left(nN\right)$. Given these values, the Newton-Raphson estimation is performed for each one of the n items. This takes $\mathcal{O}\left(nIK\right)$ time in total, where I is the number of iterations for the estimation of one item. Therefore, the total running time of item-parameters estimation is $\mathcal{O}\left(N^2K + nN + nIK\right)$. Note that the N^2K complexity can be reduced to linear using heuristics-based clustering, though the optimality is not guaranteed. Moreover, since items are independent of each other, the Newton-Raphson estimation of each item can be done in parallel, which makes the computation much more efficient than the theoretical complexity.

2) The EM *algorithm for item-parameter estimation:* Here, the goal is again to find $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ to maximize $\mathbf{P}\left(\mathbf{R} \mid \vec{\xi}\right)$. The only difference is that the elements of vector $\vec{\theta}$ are unknown. We tackle this problem using an *Expectation-Maximization* (EM) procedure.

Expectation Step: In this step, we calculate the *expected grouping* of users using previously estimated $\vec{\xi}$. In other words, for $1 \leq i \leq n$ and $1 \leq g \leq K$, we compute $\mathbf{E}\left[f_g\right]$ and $\mathbf{E}\left[r_{ig}\right]$ as follows:

$$\mathbf{E}\left[f_g\right] = \overline{f}_g = \sum_{j=1}^{N} \mathbf{P}\left(\theta_g \mid \mathbf{R}^j, \vec{\xi}\right) \text{ and} \quad (6)$$

$$\mathbf{E}\left[r_{ig}\right] = \overline{r}_{ig} = \sum_{j=1}^{N} \mathbf{P}\left(\theta_g \mid \mathbf{R}^j, \vec{\xi}\right) \times \mathbf{R}\left(i, j\right). \quad (7)$$

The computation relies on the *posterior probability distribution* of a user's attitude $\mathbf{P}\left(\theta_g \mid \mathbf{R}^j, \vec{\xi}\right)$. Assume for now that we know how to compute these probabilities. It is easy to observe that the membership of a user in a group is probabilistic. That is, every individual belongs to every group with some probability; the sum of these membership probabilities is equal to 1.

Maximization Step: Knowing the values of $\overline{f}_g$ and $\overline{r}_{ig}$ for all groups and all items allows us to compute a new estimate of $\vec{\xi}$ by invoking the Newton-Raphson item-parameters estimation procedure (NR_Item_Estimation) described in Section V-B1.

The pseudocode for the EM algorithm is given in Algorithm 1. Every iteration of the algorithm consists of an *Expectation* and a *Maximization* step.

The Posterior Probability of Attitudes: By the definition of probability, this posterior probability is:

$$\mathbf{P}\left(\theta_j \mid \mathbf{R}^j, \vec{\xi}\right) = \frac{\mathbf{P}\left(\mathbf{R}^j \mid \theta_j, \vec{\xi}\right) \mathbf{g}\left(\theta_j\right)}{\int \mathbf{P}\left(\mathbf{R}^j \mid \theta_j, \vec{\xi}\right) \mathbf{g}\left(\theta_j\right) d\theta_j}. \quad (8)$$

Algorithm 1 The EM algorithm for estimating item parameters $\xi_i = (\alpha_i, \beta_i)$ for all items $i \in \{1, \ldots, n\}$.

```
Input: Response matrix $\mathbf{R}$ and the number $K$ of user groups. Users
in the same group have the same attitude.
Output: Item parameters $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n)$, $\vec{\beta} = (\beta_1, \ldots, \beta_n)$.
for $i = 1$ to $n$ do
    $\alpha_i \leftarrow$initial_values
    $\beta_i \leftarrow$initial_values
    $\xi_i \leftarrow (\alpha_i, \beta_i)$
$\vec{\xi} \leftarrow (\xi_1, \ldots, \xi_n)$
repeat
  // Expectation step
    for $g = 1$ to $K$ do
        Sample $\theta_g$ on the ability scale
        Compute $\overline{f}_g$ using Equation (6)
        for $i = 1$ to $n$ do
            Compute $\overline{r}_{ig}$ using Equation (7).
  // Maximization step
    for $i = 1$ to $n$ do
        $(\alpha_i, \beta_i) \leftarrow$NR_Item_Estimation$\left(\mathbf{R}_i, \left\{\overline{f}_g, \overline{r}_{ig}, \theta_g\right\}_{g=1}^{K}\right)$
        $\xi_i \leftarrow (\alpha_i, \beta_i)$
until convergence
```

Function $\mathbf{g}\left(\theta_j\right)$ is the probability density function of attitudes in the population of users. It is used to model our prior knowledge about user attitudes and its called the *prior distribution* of users attitude. Following standard conventions [5], we assume that the prior distribution $\mathbf{g}(.)$ is Gaussian and is the same for all users. Our results indicate that this prior fits the data well.

The term $\mathbf{P}\left(\mathbf{R}^j \mid \theta_j, \vec{\xi}\right)$ in the numerator is the likelihood of the vector of observations $\mathbf{R}^j$ given items' parameters and user j's attitude. This term can be computed using the standard likelihood function $\mathbf{P}\left(\mathbf{R}^j \mid \theta_j, \vec{\xi}\right) = \prod_{i=1}^{n} P_{ij}^{\mathbf{R}(i,j)} \left(1 - P_{ij}\right)^{1-\mathbf{R}(i,j)}$.

The evaluation of the posterior probability of every attitude θ_j requires the evaluation of an integral. We bypass this problem as follows: Since we assume the existence of K groups, we only need to sample K points $X_1, \ldots X_K$ on the ability scale. Each of these points serves as the common attitude of a user group. For each $t \in \{1, \ldots, K\}$ we compute $\mathbf{g}\left(X_t\right)$, the density of the attitude function at attitude value X_t. Then, we let $A\left(X_t\right)$ be the *area* of the rectangle defined by the points $(X_t - 0.5, 0)$, $(X_t + 0.5, 0)$, $(X_t - 0.5, \mathbf{g}\left(X_t\right))$ and $(X_t + 0.5, \mathbf{g}\left(X_t\right))$. Then, the $A\left(X_t\right)$ values are normalized such that $\sum_{t=1}^{K} A\left(X_t\right) = 1$. In that way, we can obtain the posterior probabilities of X_t

$$\mathbf{P}\left(X_t \mid \mathbf{R}^j, \vec{\xi}\right) = \frac{\mathbf{P}\left(\mathbf{R}^j \mid X_t, \vec{\xi}\right) A\left(X_t\right)}{\sum_{t=1}^{K} \mathbf{P}\left(\mathbf{R}^j \mid X_t, \vec{\xi}\right) A\left(X_t\right)}. \quad (9)$$

Discussion: The estimation of privacy score using the IRT model requires as input the number of groups of users K. In our implementation we follow standard conventions [5] and set $K = 10$. However, we have found that other values of K fit the data as well. The estimation of the "correct" number of groups is an interesting model-selection problem for IRT models, which is not the focus of this work.

The running time of the `EM` algorithm is $\mathcal{O}\left(I_R\left(T_{\text{Exp}}+T_{\text{Max}}\right)\right)$, where I_R is the number of iterations of the `repeat` statement, and T_{Exp} and T_{Max} the running times of the `Expectation` and the `Maximization` steps respectively. Lines 9 and 11 require $\mathcal{O}\left(Nn\right)$ time each. Therefore, the total time of the expectation step is $T_{\text{Exp}} = \mathcal{O}\left(KNn^2\right)$. From the preceding discussion in Section V-B1 we know that $T_{\text{Max}} = \mathcal{O}\left(nIK\right)$, where I is the number of iterations of Equation (5). Again, Steps 12, 13, 14 can be done in parallel due to the independence assumption of items.

C. IRT-based computation of visibility

The computation of visibility requires the evaluation of $P_{ij} = \text{Prob}\left\{\mathbf{R}\left(i,j\right)=1\right\}$ given in Equation (2). Apparently if vectors $\vec{\theta} = (\theta_1,\ldots,\theta_N)$, $\vec{\alpha} = (\alpha_1,\ldots,\alpha_n)$ and $\vec{\beta} = (\beta_1,\ldots,\beta_n)$ are known, then computing P_{ij}, for every i and j, is trivial.

Here, we describe the `NR_Attitude_Estimation` algorithm, which is a Newton-Raphson procedure for computing the attitudes $\vec{\theta}$ of individuals using the item parameters $\vec{\alpha} = (\alpha_1,\ldots,\alpha_n)$ and $\vec{\beta} = (\beta_1,\ldots,\beta_n)$. These item parameters could be given as input or they can be computed using the `EM` algorithm (Algorithm 1). For each individual j, the `NR_Attitude_Estimation` computes θ_j that maximizes likelihood $\prod_{i=1}^{n} P_{ij}^{\mathbf{R}(i,j)}\left(1-P_{ij}\right)^{1-\mathbf{R}(i,j)}$, or the corresponding log-likelihood

$$L = \sum_{i=1}^{n}\left[\mathbf{R}\left(i,j\right)\log P_{ij} + \left(1-\mathbf{R}\left(i,j\right)\right)\log\left(1-P_{ij}\right)\right]. \quad (10)$$

Since $\vec{\alpha}$ and $\vec{\beta}$ are part of the input, the only variable to maximize over is θ_j. Using Newton-Raphson method, the estimate $\widehat{\theta}_j$ at iteration $(t+1)$ is computed using the estimate at iteration t as follows:

$$\left[\widehat{\theta}_j\right]_{t+1} = \left[\widehat{\theta}_j\right]_t - \left[\frac{\partial^2 L}{\partial\theta_j^2}\right]_t^{-1}\left[\frac{\partial L}{\partial\theta_j}\right]_t. \quad (11)$$

Discussion: For I iterations of the Newton-Raphson method, the running time for estimating a single user's attitude θ_j is $\mathcal{O}\left(nI\right)$. Due to the independence of users, each user's attitude is estimated separately; thus the estimation for N users requires $\mathcal{O}\left(NnI\right)$ time. Once again, this computation can be parallelized due to the independence assumption of users.

D. Putting it All Together

The sensitivity β_i computed in Section V-B and the visibility P_{ij} computed in Section V-C can be applied to Equation (1) to compute the privacy score of a user.

The advantages of the IRT framework can be summarized as follows: 1) The quantities IRT computes, *i.e.*, sensitivity, attitude and visibility, have an intuitive interpretation. For example, the sensitivity of information can be used to send early alerts to users when the sensitivities of their shared profile items are out of the comfortable region. 2) Due to the independence assumptions, many of the computations can be parallelized, which makes the computation very efficient in practice. 3) As our experiments will demonstrate later, the probabilistic model defined by IRT in Equation (2) can be viewed as a generative model, and it fits the real response data very well in terms of χ^2 goodness-of-fit test. 4) Most importantly, the estimates obtained from IRT framework satisfy the *group invariance* property. We will further discuss this property in the experimental section. At an intuitive level, this property means that the privacy scores computed across different social networks are comparable.

VI. Polytomous settings

In this section, we show how the definitions and methods described before can be extended to handle polytomous response matrices. Recall that in polytomous matrices, every entry $\mathbf{R}\left(i,j\right) = k$ with $k \in \{0,1,\ldots,\ell\}$. The smaller the value of $\mathbf{R}\left(i,j\right)$, the more conservative the privacy setting of user j with respect to profile item i. The definitions of sensitivity and visibility of items in the polytomous case generalize as follows.

Definition 1. *The sensitivity of item $i \in \{1,\ldots,n\}$ with respect to privacy level $k \in \{0,\ldots,\ell\}$, is denoted by β_{ik}. Function β_{ik} is monotonically increasing with respect to k; the larger the privacy level k picked for item i the higher its sensitivity.*

Similarly, the visibility of an item becomes a function of its privacy level.

Definition 2. *The visibility of item i that belongs to user j at level k is denoted by* $\text{V}\left(i,j,k\right)$. *The* observed visibility *is computed as* $\text{V}\left(i,j,k\right) = \mathbf{I}_{(\mathbf{R}(i,j)=k)} \times k$. *The* true visibility *is computed as* $\text{V}\left(i,j,k\right) = P_{ijk} \times k$*, where* $P_{ijk} = \textit{Prob}\left\{\mathbf{R}\left(i,j\right)=k\right\}$.

Given Definitions 1 and 2 we compute the privacy score of user j using the following generalization of Equation (1):

$$\text{Pr}\left(j\right) = \sum_{i=1}^{n}\sum_{k=0}^{\ell}\beta_{ik}\times\text{V}\left(i,j,k\right). \quad (12)$$

Again, in order to keep our framework more general, in the following sections, we will discuss true rather than observed visibility for the polytomous case.

A. IRT-based privacy score: polytomous case

Computing the privacy score in this case boils down to a transformation of the polytomous response matrix $\mathbf{R}$ into $(\ell+1)$ dichotomous response matrices $\mathbf{R}_0^*, \mathbf{R}_1^*, \ldots, \mathbf{R}_\ell^*$.

Each matrix $\mathbf{R}_k^*$, $k \in \{0, 1, \ldots, \ell\}$, is constructed so that $\mathbf{R}_k^*(i,j) = 1$ if $\mathbf{R}(i,j) \geq k$, and $\mathbf{R}_k^*(i,j) = 0$ otherwise. Let P_{ijk}^* be the probability of setting $\mathbf{R}_k^*(i,j) = 1$, *i.e.*, $P_{ijk}^* = \text{Prob}\{\mathbf{R}_k^*(i,j) = 1\} = \text{Prob}\{\mathbf{R}(i,j) \geq k\}$. When $k = 0$, matrix $\mathbf{R}_{ik}^*$ has all its entries equal to one, we have that $P_{ijk}^* = 1$ for all users. When $k \in \{1, \ldots, \ell\}$, P_{ijk}^* is given as in Equation (2). That is,

$$P_{ijk}^* = \frac{1}{1 + e^{-\alpha_{ik}^*\left(\theta_j - \beta_{ik}^*\right)}}. \tag{13}$$

By construction, for every $k', k \in \{1, \ldots, \ell\}$ and $k' < k$ we have that matrix $\mathbf{R}_k^*$ contains only a subset of the 1-entries appearing in matrix $\mathbf{R}_{k'}^*$. Therefore, $P_{ijk'}^* \geq P_{ijk}^*$, and ICC curves ($P_{ijk}^*$) of the same profile item i at different privacy levels $k \in \{1, \ldots, \ell\}$ *do not cross*. This observation results in the following corollary.

Corollary 1. *For items i and privacy levels $k \in \{1, \ldots, \ell\}$ we have that: $\beta_{i1}^* < \ldots < \beta_{ik}^* < \ldots < \beta_{i\ell}^*$. Moreover, since curves P_{ijk}^* do not cross, we also have that $\alpha_{i1}^* = \ldots = \alpha_{ik}^* = \ldots = \alpha_{i\ell}^* = \alpha_i^*$. For $k = 0$, $P_{ijk}^* = 1$, α_{i0}^* and β_{i0}^* are not defined.*

The computation of privacy score in the polytomous case, however, requires computing β_{ik} and $P_{ijk} = \text{Prob}\{\mathbf{R}(i,j) = k\}$ (see Definition (2) and Equation (12)). These parameters are different from β_{ik}^* and P_{ijk}^* since the latter are defined on dichotomous matrices. Now the question is, if we can estimate β_{ik}^* and P_{ijk}^*, how to transform them to β_{ik} and P_{ijk}.

Fortunately, since by definition P_{ijk}^* is the cumulative probability $P_{ijk}^* = \sum_{k'=k}^{\ell} P_{ijk}$, we have that

$$P_{ijk} = \begin{cases} P_{ijk}^* - P_{ij(k+1)}^* & \text{when } k \in \{0, \ldots, \ell - 1\}; \\ P_{ijk}^* & \text{when } k = \ell. \end{cases} \tag{14}$$

Also, by [2], we also have the following proposition for β_{ik}.

Proposition 1. *([2]) For $k \in \{1, \ldots, \ell - 1\}$ it holds that $\beta_{ik} = \frac{\beta_{ik}^* + \beta_{i(k+1)}^*}{2}$. Also, $\beta_{i0} = \beta_{i1}^*$ and $\beta_{i\ell} = \beta_{i\ell}^*$.*

From Proposition 1 and Corollary 1 we have the following.

Corollary 2. *For $k \in \{0, \ldots, \ell\}$ it holds that $\beta_{i0} < \beta_{i1} < \ldots < \beta_{i\ell}$.*

The above corollary verifies our intuition that the sensitivity of an item is a monotonically increasing function of the privacy level k.

Estimating the parameters: The estimation of the parameters $(\beta_{i1}^*, \ldots, \beta_{i\ell}^*, \alpha_i^*, \theta_j)$ is done using again an iterative Newton-Raphson procedure, which is pretty much the same as we did in Section V. Once these parameters are calculated, it is easy to compute P_{ijk}^*, P_{ijk} and β_{ik}. The overall privacy score is then computed by applying sensitivity values β_{ik} and visibility values $P_{ijk} \times k$ to Equation (12). We refer to the score thus obtained as the **Pr_IRT** score. The distinction between polytomous and dichotomous IRT scores becomes clear from the context.

VII. Naive privacy-score computation

In this section we describe a simple way of computing the privacy score of a user. We call this approach Naive and it serves as a baseline methodology for computing privacy scores. We also demonstrate some of its disadvantages.

Naive computation of sensitivity: Intuitively, the higher the sensitivity of an item i, the less number of people who are willing to disclose it. So, if $|\mathbf{R}_i|$ denotes the number of users who set $\mathbf{R}(i,j) = 1$, then the sensitivity β_i for dichotomous matrices can be computed as the proportion of users that are reluctant to disclose item i. That is,

$$\beta_i = \frac{N - |\mathbf{R}_i|}{N}. \tag{15}$$

The higher the value of β_i, the more sensitive item i.

For the polytomous case, the above equation generalizes as follows:

$$\beta_{ik}^* = \frac{N - \sum_{j=1}^{N} \mathbf{I}_{(\mathbf{R}(i,j) \geq k)}}{N}. \tag{16}$$

Note that the β_{ik} values are then computed from β_{ik}^* following Proposition 1.

Naive computation of visibility: The computation of visibility in the dichotomous case requires an estimate of the probability $P_{ij} = \text{Prob}\{\mathbf{R}(i,j) = 1\}$. Assuming independence between items and individuals, we can compute P_{ij} to be the product of the probability of an 1 in row $\mathbf{R}_i$ times the probability of an 1 in column $\mathbf{R}^j$. That is, if $\left|\mathbf{R}^j\right|$ is the number of items for which j sets $\mathbf{R}(i,j) = 1$, we have

$$P_{ij} = \frac{|\mathbf{R}_i|}{N} \times \frac{\left|\mathbf{R}^j\right|}{n}. \tag{17}$$

Probability P_{ij} is higher for less sensitive items and for users that have the tendency/attitude to disclose lots of their profile items.

The visibility in the polytomous case requires the computation of probability $P_{ijk} = \text{Prob}\{\mathbf{R}(i,j) = k\}$. By assuming independence between items and users, this probability can be computed as follows:

$$P_{ijk} = \frac{\sum_{j=1}^{N} \mathbf{I}_{(\mathbf{R}(i,j) = k)}}{N} \times \frac{\sum_{i=1}^{n} \mathbf{I}_{(\mathbf{R}(i,j) = k)}}{n}. \tag{18}$$

The Naive computation of privacy score requires applying Equations (15) and (17) to Equation (1), or Equations (16) and (18) to Equation (12). We refer to the privacy-risk score computed in this way as the **Pr_Naive** score.

Discussion: The Naive computation can be done efficiently in $\mathcal{O}(Nn)$ time. But the disadvantage is that the sensitivity values obtained are significantly biased by the user population contained in $\mathbf{R}$. If the users happen to be quite conservative and they rarely share anything, then the estimated sensitivity values can be very high, otherwise the values can be very low if the users are very extrovert. Moreover, as we will show in the experimental section, the probability model defined by Equation (17) and (18), though simple and intuitive, fails to fit the real-world response matrices $\mathbf{R}$ (in terms of χ^2 goodness-of-fit).

VIII. Experiments

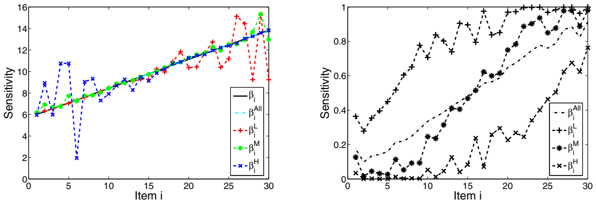
The purpose of the experimental section is to illustrate the properties of the different methods for computing users' privacy scores and pinpoint their advantages and disadvantages. From the data-analysis point of view, our experiments with real data show interesting facts about the users' behavior.

A. *Datasets*

We start by giving a brief description of the synthetic and real-world datasets we used for our experiments.

Dichotomous Synthetic dataset: This dataset consists of a dichotomous $n \times N$ response matrix $\mathbf{R}_S$ where the rows correspond to items and the columns correspond to users. The response matrix $\mathbf{R}_S$ is generated as follows: for each item i, of a total of $n = 30$ items, we pick parameters α_i and β_i uniformly at random from intervals $(0, 2)$ and $[6, 14]$ respectively. We assume that the items are sorted based on their β_i values, i.e., $\beta_1 < \beta_2 < \ldots < \beta_n$. Next, $K = 30$ different attitude values are picked uniformly at random from the real interval $[6, 14]$. Each such attitude value θ_g is associated with a group of 200 users (all 200 users in a group have attitude θ_g). Let the groups be sorted so that $\theta_1 < \theta_2 < \cdots < \theta_K$. For every group F_g, user $j \in F_g$ and item i, we set $\mathbf{R}_s(i, j) = 1$ with probability $\text{Prob}\{\mathbf{R}(i, j) = 1\} = 1/\left(1 + e^{-\alpha_i(\theta_g - \beta_i)}\right)$.

Survey dataset: This dataset consists of the data we collected by contacting an online survey [7]. The goal of the survey is to collect users' information-sharing preferences. Given a list of profile items that span a large spectrum of one's personal life (*e.g.*, name, gender, birthday, political views, interests, address, phone number, degree, job, etc.), the users are asked to specify the extent they want to share each item with others. The privacy levels a user can allocate to items are $\{0, 1, 2, 3, 4\}$. Value **0** means that a user wants to share this item with *no one*; **1** means that he wants to share it with *some of his immediate friends*, **2** with *all of his immediate friends*, **3** with *all immediate friends and friends of friends*, and **4** with *everyone*. This setting simulates most of the privacy-setting options used in real online social networks. Along with users' privacy settings, we also collect information about their location, educational background, age etc. The survey spans 49 profile items. We have received 153 complete responses from 18 countries/political regions. Among the participants, 53.3% are male and 46.7% are female, 75.4% are in the age of 23 to 39, 91.6% hold a college degree or higher, and 76.0% spend 4 hours or more everyday surfing online.



(a) IRT model

(b) Naive model

Figure 2. Testing the group-invariance property of item parameter estimation using IRT (Figure 2(a)) and Naive (Figure 2(b)) models.

From the **Survey** dataset we construct a polytomous response matrix $\mathbf{R}$ (with $\ell = 4$). This matrix contains the privacy levels picked by the 153 respondents for each one of the 49 items. We also construct four dichotomous matrices $\mathbf{R}_k^*$ with $k = \{1, 2, 3, 4\}$ as follows: $\mathbf{R}_k^*(i, j) = 1$ if $\mathbf{R}(i, j) \geq k$, and 0 otherwise.

B. *Experiments with the **Dichotomous Synthetic** data*

The goal of the experiments in this section is to demonstrate the *group invariance* property of the IRT model using the **Dichotomous Synthetic** dataset.

We conduct the experiment as follows: first, we cluster the 6000 users into 3 groups $F_L = \cup_{g=1\ldots10} F_g$, $F_M = \cup_{g=11\ldots20} F_g$ and $F_H = \cup_{g=21\ldots30} F_g$. That is, the first cluster consists of all users in the 10 lowest-attitude groups $F_1, \ldots, F_{10}$, the second cluster consists of all users in the 10 medium-attitude groups and the third one consists of all users in the 10 highest-attitude groups. Given users' attitudes assigned in the data-generation, we estimate three sets of item parameters $\xi_i^L = \left(\alpha_i^L, \beta_i^L\right)$, $\xi_i^M = \left(\alpha_i^M, \beta_i^M\right)$ and $\xi_i^H = \left(\alpha_i^H, \beta_i^H\right)$ for every item i. The estimation is done using Equation (5) with input response matrix that only contains the columns of $\mathbf{R}_S$ associated with the users in F_L, F_M and F_H, respectively. We use Equation (5) to compute estimates $\xi_i^{all} = \left(\alpha_i^{all}, \beta_i^{all}\right)$ using the whole response matrix $\mathbf{R}_S$.

Figure 2(a) shows the estimated sensitivity values of the items. Since the data was generated using IRT model, the true sensitivity parameters β_i for each item are also known (and plotted). The x-axis of the figure shows the different items sorted in increasing order of their true β_i values. It can be seen that for the majority of the items, the estimated sensitivity values β_i^L, β_i^M, β_i^H and β_i^{all} are all very close to the true β_i value. This indicates one of the interesting features of IRT that item parameters are not dependent upon the attitude level of the users responding to the item. Thus, the item parameters are what is known as *group invariance*. The validity of this property is demonstrated in Frank Baker's book [2] and online tutorial: http://echo.edres.org:

[7] http://www.surveymonkey.com/s.aspx?sm=G2qidwUNtRKCgNIbLP2bBg_3d_3d

8080/irt/baker/chapter3.pdf. At an intuitive level, since the same item was administrated to all groups, each of the three parameter estimation processes was dealing with a segment of the same underlying item characteristic curve (see Figure 1). Consequently, the item parameters yielded by the three estimations should be identical.

It should be noted that even though the item parameters are group invariant, this does not mean that in practice values of the same item parameter estimated from different groups of users will always be exactly the same. The obtained values will be subject to variation due to group size, the goodness-of-fit of the ICC curve to the data. Nevertheless, the estimated values should be in "the same ballpark". This explains why in Figure 2(a) there are some items for which the estimated parameters deviate from the true one more.

We repeat the same experiment for the Naive model. That is, for each item we estimate sensitivities β_i^L, β_i^M, β_i^H and β_i^{all} using the Naive approach (Section VII). Figure 2(b) shows the obtained estimates. The plot demonstrates that the Naive computation of sensitivity does not have the group-invariance property. For most of the items, sensitivity β_i^L obtained from users with low attitude levels (*i.e.*, conservative, introvert) are much higher than the β_i^{all} estimates since these users rarely share anything, whereas β_i^H obtained from users with high attitude levels (*i.e.*, careless, extrovert) are much lower than β_i^{all}.

Note that since sensitivities estimated by Naive and IRT are not in the same scale, one should consider the relative error instead of absolute error when comparing the results in Figures 2(a) and 2(b).

C. Experiments with the ***Survey*** *data*

The goal of the experiments in this section is to use the **Survey** dataset to show 1) IRT is a good model for the real-world data, whereas Naive is not; and 2) IRT model provides us an interesting estimation of the sensitivity of information being shared in online social networks.

1) Testing χ^2 goodness-of-fit: We start by illustrating that the IRT model fits the real-world data very well, whereas the Naive model does not. For that we use χ^2 *goodness-of-fit test*, a commonly-used test for accepting or rejecting the *null hypothesis* that a data sample comes from a specific distribution. Our input data consists of dichotomous matrix $\mathbf{R}_k^*$ ($k \in \{1, 2, 3, 4\}$) constructed from **Survey** data.

First we test whether IRT model is a good model for data in each $\mathbf{R}_k^*$. We test this hypothesis as follows: first use the EM algorithm to estimate both items' parameters and users' attitudes. Then, we use an 1-dimensional dynamic-programming algorithm to group the users based on their estimated attitudes. The mean attitude of a group F_g serves as the group attitude θ_g. Next, for each matrix $\mathbf{R}_k^*$ and each item i in the matrix, we compute

$$\chi^2 = \sum_{g=1}^{K} \left(\frac{(f_g \tilde{p}_{ig} - f_g p_{ig})^2}{f_g p_{ig}} + \frac{(f_g \tilde{q}_{ig} - f_g q_{ig})^2}{f_g q_{ig}} \right).$$

In the above equation, f_g is the number of users in group F_g; p_{ig} (resp. $\tilde{p}_{ig}$) is the expected (resp. observed) proportion of users in F_g that set $\mathbf{R}_k^*(i, j) = 1$. Finally, $q_{ig} = 1 - p_{ig}$ (and $\tilde{q}_{ig} = 1 - \tilde{p}_{ig}$). For the IRT model $p_{ig} = P_i(\theta_g)$ and it is computed using Equation (2) for group attitude θ_g and item parameters estimated by EM. For IRT, the test statistic follows, approximately, a χ^2-distribution with $(K - 2)$ degrees of freedom since there are 2 estimated parameters.

For testing whether the responses in $\mathbf{R}_k^*$ can be described by the Naive model we follow a similar procedure. First we compute, for each user, the proportion of items that the user sets equal to 1 in $\mathbf{R}_k^*$. This value serves as the user's "pseudo-attitude". Then we construct K groups of users $F_1, \ldots, F_K$, using an 1-dimensional dynamic-programming algorithm based on these attitude values. Given this grouping, the χ^2 statistic is computed again. The only difference here is that

$$p_{ig} = \left(\frac{|\mathbf{R}_{k_i}^*|}{N} \right) \times [\frac{1}{f_g} \sum_{j \in F_g} \frac{|\mathbf{R}_k^{*^j}|}{n}], \qquad (19)$$

where $|\mathbf{R}_{k_i}^*|$ denotes the number of users who shared item i in $\mathbf{R}_k^*$, and $|\mathbf{R}_k^{*^j}|$ denotes the number of items being shared by a user j in $\mathbf{R}_k^*$. For Naive, the test statistic approximately follows a χ^2-distribution with $(K - 1)$ degrees of freedom.

Table I shows the number of items for which the null hypothesis that their responses follow the IRT or Naive model is rejected. We show results for dichotomous matrices $\mathbf{R}_1^*$, $\mathbf{R}_2^*$, $\mathbf{R}_3^*$, and $\mathbf{R}_4^*$ and $K = \{6, 8, 10, 12, 14\}$. In all cases, the null hypothesis that the responses follow the Naive model were rejected for all 49 items. On the other hand, the null hypothesis that responses follow the IRT model was rejected only for small number of items in all configurations. This gives strong evidence that the IRT model fits the real data better. All results reported here are for confidence level .95.

	$\mathbf{R}_1^*$	$\mathbf{R}_2^*$	$\mathbf{R}_3^*$	$\mathbf{R}_4^*$	
	IRT				Naive
K=6	4	3	6	11	49
K=8	4	3	4	8	49
K=10	5	5	7	8	49
K=12	5	3	5	7	49
K=14	5	3	3	7	49

Table I
$\mathbf{R}_2^*$ DATA – χ^2-GOODNESS-OF-FIT TESTS: THE NUMBER OF REJECTED HYPOTHESES (OUT OF A TOTAL OF 49) WITH RESPECT TO THE NUMBER OF GROUPS K.

2) Sensitivity of profile items: In Figure 3 we visualize, using a tag cloud, the sensitivity of the profile items used in our survey. The evaluation of sensitivity values is done using the EM algorithm (Algorithm 1) with input the dichotomous response matrix $\mathbf{R}_2^*$. The larger the fonts used to represent a profile item in the tag cloud, the higher its estimated sensitivity value. It is easily observed that *Mother's Maiden Name* is the most sensitive item, while one's *Gender*, which locates just right above the letter "h" of "Mother" has the lowest sensitivity; too small to be visually identified.

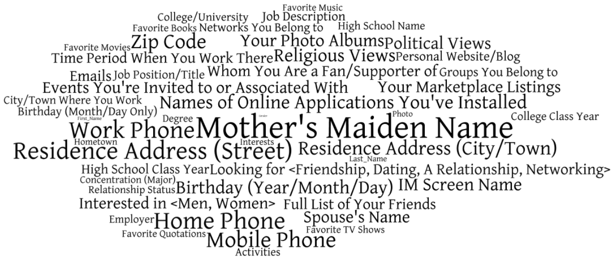


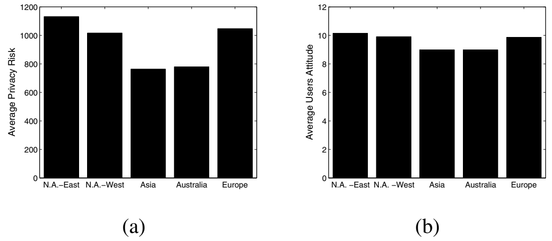
Figure 3. Sensitivity of the profile items computed using IRT model with input the dichotomous matrix $\mathbf{R}_2^*$. Larger fonts mean higher sensitivity.

3) Geographic distribution of privacy scores: Here we present some interesting findings we get by further analyzing the **Survey** dataset. We compute the the privacy scores of the 153 respondents using the polytomous IRT-based computations (Section VI-A).

After evaluating the privacy scores of individuals using as input the whole response matrix $\mathbf{R}$, we group the respondents based on their geographic location. Figures 4(a) and 4(b) show the average values of the users' **Pr_IRT** scores and attitudes per location. The results indicate that people from North America and Europe have higher privacy scores than people from Asia and Australia. The privacy scores and the attitude values are highly correlated. This experimental finding indicates that people from North America and Europe are more comfortable to reveal personal information on the social networks they participate. This can be either a result of inherent attitude or social pressure. Since online social-networking is more widespread in these regions, one can assume that people in North America and Europe succumb to the social pressure to reveal things about themselves online in order to appear "cool" and become popular.

IX. Conclusions

We have presented models and algorithms for computing the privacy score of users in online social networks. Our methodology takes into account the privacy settings of users with respect to their profile items and has its mathematical underpinnings in Item Response Theory. We have described the bases of our approach and presented a set of experiments on synthetic and real data that highlight the properties of our model and the current trends in users behavior. Our framework gives a new way of dealing with privacy in social networks. Although it is by no means perfect, it opens door to various possibilities of privacy management that will make our online environment more comfortable.



(a)

(b)

Figure 4. **Survey** data: average privacy scores (**Pr_IRT**) (Figure 4(a)) and average users attitudes (Figure 4(b)) per geographic region.

Acknowledgment

The authors would like to thank Tyrone Grandison and Michael Maximilien for their valuable comments and suggestions on this work.

References

[1] L. Backstrom, C. Dwork, and J. M. Kleinberg. Wherefore art thou R3579X?: Anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*, pages 181–190, Alberta, Canada, May 2007.

[2] F. B. Baker and S.-H. Kim. *Item Response Theory: Parameter Estimation Techniques*. Marcel Dekkerm, Inc., 2004.

[3] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis. Resisting structural re-identification in anonymized social networks. *Proceedings of the VLDB Endowment*, 1(1):102–114, August 2008.

[4] K. Liu and E. Terzi. Towards identity anonymization on graphs. In *Proceedings of ACM SIGMOD*, pages 93–106, Vancouver, Canada, June 2008.

[5] R. Mislevy and R. Bock. PC-BILOG: Item analysis and test scoring with binary logistic models, 1986.

[6] M. Richardson, R. Agrawal, and P. Domingos. Trust management for the semantic web. In *International Semantic Web Conference*, pages 351–368, 2003.

[7] X. Ying and X. Wu. Randomizing social networks: a spectrum preserving approach. In *Proceedings of 2008 SIAM International Conference on Data Mining (SDM'08)*, pages 739–750, Atlanta, GA, April 2008.

[8] T. J. Ypma. Historical development of the Newton-Raphson method. *SIAM Rev.*, 37(4):531–551, 1995.

[9] B. Zhou and J. Pei. Preserving privacy in social networks against neighborhood attacks. In *Proceedings of the 24th International Conference on Data Engineering (ICDE'08)*, pages 506–515, Cancun, Mexico, April 2008.