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

KUN LIU
Yahoo! Labs
and
EVIMARIA TERZI
Boston University

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

In this article, we approach the privacy issues raised in online social networks from the individual users' viewpoint: we propose a framework to compute the privacy score of a user. This score indicates the user's potential risk caused by his or her participation in the network. Our definition of privacy score satisfies the following intuitive properties: the more sensitive information a user discloses, the higher his or her 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.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications—*Data mining*

General Terms: Algorithms, Experimentation, Theory

Additional Key Words and Phrases: Social networks, item-response theory, expectation maximization, maximum-likelihood estimation, information propagation

ACM Reference Format:
Liu, K. and Terzi, E. 2010. A framework for computing the privacy score of users in online social networks. ACM Trans. Knowl. Discov. Data 5, 1, Article 6 (December 2010), 30 pages.
DOI: 10.1145/1870096.1870102. http://doi.acm.org/10.1145/1870096.1870102.

A shorter version of this article appeared in the *Proceedings of the 2009 International Conference on Data Mining* (ICDM).
This work was done while K. Liu was with the IBM Almaden Research Center.

Authors' addresses: K. Liu, Yahoo! Labs, 4401 Great America Parkway, Santa Clara, CA 95054; email: kyn@yahoo-inc.com; E. Terzi (contact author), Computer Science Department, Boston University, 111 Cummington Street, Boston, MA 02215; email: evimaria@cs.bu.edu.

1. INTRODUCTION

In recent years, online social networking has moved from niche phenomenon to mass adoption. As we are writing this article, the two largest social-networking Web sites in the U.S., Facebook and MySpace, each already has more than 110 million monthly active users [Owyang 2008]. The goal of users upon entering a social network is to contact or be contacted by others, meet new friends or dates, find new jobs, receive or provide recommendations, and much more.

Liu and Maes [2005] estimated that well over a million self-descriptive personal profiles are available across different Web-based social networks in the United States. According to Leonard [2004], already in 2004, seven million people had accounts on Friendster, two million users were registered with MySpace, while a whopping sixteen million users were registered on Tickle for a chance to take a personality test. Facebook, for example, has spread to millions of users [Gross and Acquisti 2005], that span of various educational institutions, high-school, undergraduate, and graduate students, and faculty members, staff, and alumni—not to mention the vast media attention it has received.

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 experienced users who are aware of their privacy risks are sometimes willing to compromise their privacy in order to improve their digital presence in the virtual world. That is, they prefer being popular and "cool" to being conservative with respect to their privacy settings. 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., Backstrom et al. [2007]; Hay et al. [2008]; Liu and Terzi [2008]; Ying and Wu [2008]; Zhou and Pei [2008]) has mainly focused on corporate-scale privacy concerns, that is, how to share a social network has owned by an organization without revealing the identities of or sensitive relationships among the registered users. Not much attention has been given to the privacy risk for individual users posed by their information-sharing activities.

In this article, 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 or her online information-sharing behavior. With this score, we can achieve the following.

—*Privacy risk monitoring*. The score serves as an indicator of the user's potential privacy risk. The system can estimate the sensitivity of each piece of information the user has shared, and send alert to the user if the sensitivity of some information is beyond the predefined threshold.

—*Privacy setting recommendation*. The user can compare his or her privacy score with the rest of the population to know where he or she stands. In the case where the overall privacy score of a user's social graph is lower than that of the user himself or herself, the system can recommend stronger privacy settings based on information from the user's social neighbors.

- *Social study*. As a byproduct, the system can estimate the inherent attitude of each individual. This psychometric measure can help sociologists study the online behavior of users.

The overall objective of our work is to enhance public awareness of privacy, and to reduce the complexity of managing 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 (i) the *sensitivity* of the information being revealed and (ii) the *visibility* of the revealed information within the network. We develop mathematical models to estimate both the 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 propose in this article 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 scores that scale well and indicative experimental evidence of the efficacy of our framework. Our models draw inspiration from the *Item Response Theory* (IRT) [Baker and Kim 2004] and *Information Propagation* (IP) models [Kempe et al. 2003].
- *Overview of our framework*. For a social-network user, j, we compute the *privacy score* as a combination of the partial privacy scores of each one of his or her profile items, for example, the user's real name, email, hometown, mobile-phone number, relationship to status, sexual orientation, IM screen name, etc. The contribution of each profile item to the total privacy score depends on the sensitivity of the item and the visibility it gets due to j's privacy settings and j's position in the network.

Here, we assume that all N users specify their privacy settings for the same n profile items. These settings are stored in an $n \times N$ *response matrix* $\mathbf{R}$. The profile setting of user j for item i, $\mathbf{R}(i, j)$, is an integer value that determines how willing j is to disclose information about i. The higher the value, the more willing j is to disclose information about item i. In general, large values in $\mathbf{R}$ imply higher visibility. On the other hand, small values in the privacy settings of an item are an indication of high sensitivity; it is the highly sensitive items that most people try to protect. Therefore, the privacy settings of users for their profile items stored in the response matrix $\mathbf{R}$ have lots of valuable information about users' privacy behavior. Our first approach uses exactly this information to compute the privacy score of users. We do so by employing notions from Item Response Theory (IRT) [Baker and Kim 2004]. The position of every user in the social network also affects his or her privacy score. The visibility setting of the profile items is enhanced (or silenced) depending on the user's role in the network. For example, the privacy risk of a completely isolated individual is much lower than the privacy risk of a popular individual, even if both have the same privacy settings in their profiles. In our extended version of privacy-score computation, we take into account the social-network structure and use models and algorithms from information-propagation and viral marketing studies [Kempe et al. 2003].

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

Although the privacy scores computed by a single method are all comparable (i.e., they are on the same scale), the scale across different methods varies. Adjusting the scales of measures that have totally different ranges can sometimes be tricky, and crude normalization can lead to misinterpretations of the results. In this article, we do not adjust the scales of different scores. Instead, we emphasize the properties of these scores and the ranking of users with respect to their privacy scores.

—*Organization of the material*. After the presentation of the related work in Section 2 and the description of the notational conventions in Section 3, we present our definitions of privacy score and present algorithms for computing it. Our initial privacy score definition ignores the structure of the social network; its computation is only based on the privacy settings users associate with their profile items. The models and algorithmic solutions associated with this definition of privacy score are given in Sections 4, 5, 6, and 7. We extend our definition of privacy score to take into account the social-network structure in Section 8. Experimental comparison of our methods is given in Section 9. We conclude the article in Section 10.

2. RELATED WORK

To the best of our knowledge, we are the first to present a framework that formally quantifies the privacy score of online social-network users. None of the previous work on privacy-preserving social-network analysis [Backstrom et al. 2007; Hay et al. 2008; Liu and Terzi 2008; Ying and Wu 2008; Zhou and Pei 2008] has addressed privacy concerns from this perspective. Past work has mostly considered anonymization methods and threats to users' privacy once an anonymized social network is released. What we consider as the most relevant work is that 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 presence. The QDOS score is determined by four factors: (1) popularity, that is, who and how many people know you; (2) impact, that is, the extent to which people are influenced by what you say; (3) activity, that is, what you do online; and (4) individuality, that is, how easily you can be located online. Although QDOS can be potentially used to measure one's privacy risk, the primary purpose of this system as of today is the opposite; it encourages people to enhance their digital presence. More importantly, QDOS uses a different mathematical model based on spectral analysis of the input social network. Our model, on the other hand, exploits item response theory and information-propagation models.

—*Credit score*. A credit score is used to estimate the likelihood that a person will default on a loan. The most famous one, the FICO score,[2] was originally developed by Fair Isaac Corporation in 1956. Nowadays, this ubiquitous three-digit number is used to evaluate the creditworthiness of a person. The credit score is different from our privacy 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 much one a member of a group is trusted by the others. There is a large body of applications and research on this topic, see for example, eBay's sellers and buyers rating system, trust management for the Semantic Web [Richardson et al. 2003], etc. 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.

Ahmad [2006] described a method for managing the release of private information. When a request is received, the information provider calculates a score to serve as a confidence level for authentication and authorization associated with the request. The provider releases the information only when the score is above a predefined threshold.

—*Identity score*. An identity score[3] is used for tagging and verifying the legitimacy of a person's public identity. It was originally developed by financial-service firms to measure the fraud risk of new customers. Our privacy score is different from an identity score since it serves a different purpose.

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

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

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

3. 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. Users are connected through links that correspond to the edges of $\mathcal{G}$. In principle, the links are unweighted and undirected. However, for generality, we assume that $\mathcal{G}$ is directed and we have converted undirected networks into directed ones by adding two directed edges $(j \rightarrow j')$ and $(j' \rightarrow j')$ for every input undirected edge (j, j'). 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 privacy levels picked by all N users for the n profile items are stored in an $n \times N$ *response matrix* $\mathbf{R}$. The rows of $\mathbf{R}$ correspond to profile items and the columns correspond to users. We use $\mathbf{R}(i, j)$ to refer to the entry in the ith row and jth column of $\mathbf{R}$; $\mathbf{R}(i, j)$ refers to the privacy setting of user j for item i. If the entries of the response matrix $\mathbf{R}$ are restricted to take values in $\{0, 1\}$, we say that $\mathbf{R}$ is a *dichotomous* response matrix. If entries in $\mathbf{R}$ take any nonnegative integer values in $\{0, 1, \ldots, \ell\}$, we say that matrix $\mathbf{R}$ is a *polytomous* response matrix.

In a dichotomous response matrix $\mathbf{R}$, $\mathbf{R}(i, j) = 1$ means that user j has made the information associated with profile item i publicly available. If user j has kept information related to item i private, then $\mathbf{R}(i, j) = 0$. The interpretation of values appearing in polytomous response matrices is similar: $\mathbf{R}(i, j) = 0$ means that user j keeps profile item i private; $\mathbf{R}(i, j) = 1$ means that j discloses information regarding item i only to his or her immediate friends. In general, $\mathbf{R}(i, j) = k$ (with $k \in \{0, 1, \ldots, \ell\}$) means that j discloses information related to item i to users that are at most k links away in $\mathcal{G}$.

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

In most of the social media sites (e.g., Facebook, Flickr, LinkedIn, etc.), there is a response matrix where the user is asked to determine his or her choice of privacy levels for different information items. In some cases, the privacy levels are dichotomous or polytomous. It can be the case that some of the users do not set levels in all their profile items, for some of them leave the default settings. Therefore, the response matrix is always complete for every user. Recent research [Fang and LeFevre 2010] has also aimed at helping users set their personalized levels for all items, based on their selected levels at a subset of those items.

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, $P_{ij} = \text{Prob}\left\{\mathbf{R}(i, j) = 1\right\}$. In the polytomous case, we use P_{ijk} to denote the probability that user j sets $\mathbf{R}(i, j) = k$. That is, $P_{ijk} = \text{Prob}\left\{\mathbf{R}(i, j) = k\right\}$.

In order to allow the readers to build intuition, we start by defining the privacy score for dichotomous response matrices. Once this intuition is built, we extend our definitions to polytomous settings.

4. PRIVACY SCORE IN DICHOTOMOUS SETTINGS

The privacy score of a user is an indicator of his or her potential privacy risk; the higher the privacy score of a user, the higher the threat to his or her privacy. Naturally, the privacy risk of a user depends on the privacy level he or she picks for the his or her profile items. The basic premises of our definition of privacy score are the following.

—The more sensitive information a user reveals, the higher his or her privacy score.

—The more people know some piece of information about a user, the higher his or her privacy score.

The following two examples illustrate these two premises.

Example 1. Assume user j and two profile items, $i = \{$*mobile-phone number*$\}$ and $i' = \{$*employer*$\}$. $\mathbf{R}(i, j) = 1$ is a much more risky setting for j than $\mathbf{R}(i', j) = 1$; even if a large group of people knows j's employer, this cannot be as intrusive a scenario as the one where the same set of people knows js mobile-phone number.

Example 2. Assume again user j and let $i = \{$*mobile-phone number*$\}$ be a single profile item. Naturally, setting $\mathbf{R}(i, j) = 1$ is a more risky behavior than setting $\mathbf{R}(i, j) = 0$; making j's mobile phone publicly available increases j's privacy risk.

In order to capture the essence of the preceding examples, we define the privacy score of user j to be a monotonically increasing function of two parameters: the sensitivity of the profile items and the visibility these items get.

4.1 Sensitivity of a Profile Item

Examples 1 and 2 illustrate that the sensitivity of an item depends on the item itself. Therefore, we define the sensitivity of an item as follows.

Definition 1. The sensitivity of item $i \in \{1, \ldots, n\}$ is denoted by β_i and depends on the nature of the item i.

Some profile items are, by nature, more sensitive than others. In Example 1, the $\{$*mobile-phone number*$\}$ is considered more sensitive than $\{$*employer*$\}$ for the same privacy level.

4.2 Visibility of a Profile Item

The visibility of a profile item i due to j captures how known j's value for i becomes in the network; the more it spreads, the higher the item's visibility.

Naturally, visibility, denoted by $\mathrm{V}(i, j)$, depends on the value $\mathbf{R}(i, j)$, as well as on the particular user j and his or her position in the social network $\mathcal{G}$. The simplest possible definition of visibility is $\mathrm{V}(i, j) = \mathbf{I}_{(\mathbf{R}(i,j)=1)}$, 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 and user j. In general, one can assume that $\mathbf{R}$ is a sample from a probability distribution over all possible response matrices. Then the *true visibility*, or simply the visibility, is computed based on this assumption.

Definition 2. If $P_{ij} = Prob\{\mathbf{R}(i, j) = 1\}$, then the visibility is $\mathrm{V}(i, j) = P_{ij} \times 1 + (1 - P_{ij}) \times 0 = P_{ij}$.

Probability P_{ij} depends both on the item i and the user j.

4.3 Privacy Score of a User

The privacy score of individual j due to item i, denoted by $\mathrm{P{\scriptsize R}}(i, j)$, can be any combination of sensitivity and visibility. That is,

$$\mathrm{P{\scriptsize R}}(i, j) = \beta_i \bigotimes \mathrm{V}(i, j).$$

Operator $\bigotimes$ is used to represent any arbitrary combination function that respects the fact that $\mathrm{P{\scriptsize R}}(i, j)$ is monotonically increasing with both sensitivity and visibility. For simplicity, throughout our discussion we use the product operator to combine sensitivity and visibility values.

In order to evaluate the overall privacy score of user j, denoted by $\mathrm{P{\scriptsize R}}(j)$, we can combine the privacy score of j due to different items. Again, any combination function can be employed to combine the per-item privacy scores. For simplicity, we use a summation operator here. That is, we compute the privacy score of individual j as follows:

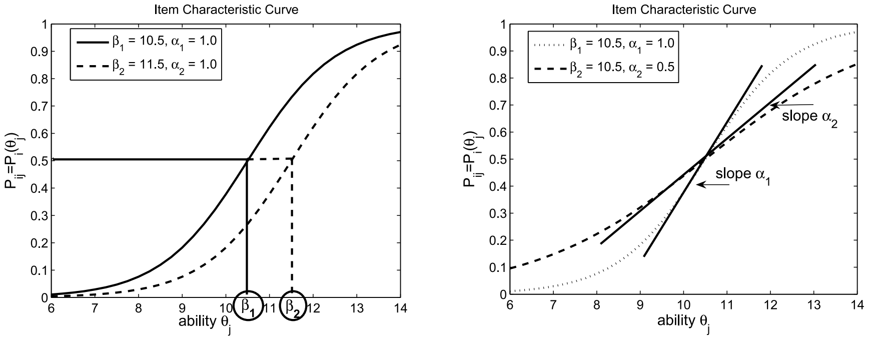
$$\mathrm{P{\scriptsize R}}(j) = \sum_{i=1}^{n} \mathrm{P{\scriptsize R}}(i, j) = \sum_{i=1}^{n} \beta_i \times \mathrm{V}(i, j). \tag{1}$$

In the above, the privacy score can be computed using either the observed visibility or the true visibility. For the rest of the discussion, we use the true visibility, and we refer to it as visibility. This is because we believe that the specific privacy settings of a user are just an instance of his or her possible settings descibed by the probability distribution $\mathrm{P{\scriptsize R}}(i, j)$.

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

5. IRT-BASED COMPUTATION OF PRIVACY SCORE: DICHOTOMOUS CASE

In this section we show how to compute the privacy score of users using concepts from Item Response Theory (IRT). We start the section by introducing some basic concepts from IRT. We then show how these concepts are applicable in our setting.



(a) $\beta_1 < \beta_2$.

(b) $\alpha_1 > \alpha_2$.

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

5.1 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 correctly answering a given question.

In this article, we consider the two-parameter IRT model. In this model, 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. Every examinee j is characterized by his or her ability level θ_j, $\theta_j \in (-\infty, \infty)$. The basic random variable of the model is the response of examinee j to a particular question q_i. If this response is marked either "correct" or "wrong" (dichotomous response), then the probability that j answers q_i *correctly* is given by

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

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[4] is called the *item characteristic curve* (ICC).

The ICCs obtained for different values of parameters $\xi_i = (\alpha_i, \beta_i)$ are given in Figures 1(a) and 1(b). These illustrations 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 compared. If an examinee's ability is higher than

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

the difficulty of the question, then he or she has a better chance to get the answer 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 $P_{ij} = P_i\left(\theta_j\right)$ 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 $\text{Prob}\left\{\mathbf{R}(i, j) = 1\right\}$ using Equation (2). However, we do not have examinees and questions; rather we 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 or her attitude θ_j quantifies how concerned j is about his or her 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 preceding mapping, parameter α_i is ignored. Naturally, the need of the two-parameter model (q_i is characterized by α_i, β_i) is questioned. One could argue that for our purposes it is enough to use the one-parameter model (q_i is only characterized by β_i), which is also known as the *Rasch model*.[5] In the Rasch model, each item is described by parameter β_i and $\alpha_i = 1$ for all $i \in \{1, \ldots, n\}$. However, as shown in Birnbaum [1968] (and discussed in Baker and Kim [2004], Chapter 5), the Rasch model is unable to distinguish users that disclose the same number of profile items but with different sensitivities. We believe that a finer-grained analysis of users attitude is necessary and this is the reason we pick the two-parameter model.

For computing the privacy score, we need to compute the sensitivity β_i for all items $i \in \{1, \ldots, n\}$ and the probabilities $P_{ij} = \text{Prob}\left\{\mathbf{R}(i, j) = 1\right\}$, 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 next three 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 inherent in IRT models: (i) independence between items; (ii) independence between users; and (iii) independence between users and items. The independence assumptions are necessary for devising a relatively simple and intuitive model. Also, they help in the design of efficient algorithms for

[5] http://en.wikipedia.org/wiki/Rasch_model

computing the privacy scores. Modeling dependencies between users or items would significantly increase the computational complexity of our methods and it would make them incapable of handling large datasets in real scenarios. Further, our experiments in Section 9 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.

5.2 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 5.2.1 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 5.2.2.

5.2.1 *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*

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

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 a different attitude per user. In practice, online social-network users form a grouping that partitions the set of users $\{1, \ldots, N\}$ into K nonoverlapping groups $\{F_1, \ldots, F_K\}$ such that $\bigcup_{g=1}^{K} F_g = \{1, \ldots, N\}$. All users within each partition have a similar "attitude." Let θ_g be the attitude of group F_g (all members of F_g share the same attitude θ_g) and $f_g = \left|F_g\right|$. 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 a grouping, the likelihood function can be written as

$$\prod_{g=1}^{K} \binom{f_g}{r_{ig}} \left[P_i\left(\theta_g\right)\right]^{r_{ig}} \left[1 - P_i\left(\theta_g\right)\right]^{f_g - r_{ig}}.$$

After ignoring the constants, the corresponding log-likelihood function is

$$L = \sum_{g=1}^{K} \left[r_g \log P_i\left(\theta_g\right) + \left(f_g - r_{ig}\right) \log\left(1 - P_i\left(\theta_g\right)\right)\right]. \tag{3}$$

The partitioning of the users into groups is made for two reasons: (1) the partitioning reduces the computational complexity of the algorithm. We only have to consider the groups of users rather than the users. (2) Partitioning of the users into groups that have the same "attitude" parameters allows us to

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

get better estimates of the parameters of the groups as well as of the sensitivity values of the items, since we have more observations per parameter value.

Our goal is now to find item parameters $\xi_i = (\alpha_i, \beta_i)$ to maximize the log-likelihood function given in Equation (3). For this we use the Newton-Raphson method [Ypma 1995]. The Newton-Raphson method is a numerical algorithm that, given partial derivatives

$$L_1 = \frac{\partial L}{\partial \alpha_i} \quad \text{and} \quad L_2 = \frac{\partial L}{\partial \beta_i}$$

and

$$L_{11} = \tfrac{\partial^2 L}{\partial \alpha_i^2}, \quad L_{22} = \tfrac{\partial^2 L}{\partial \beta_i^2}, \quad L_{12} = L_{21} = \tfrac{\partial^2 L}{\partial \alpha_i \beta_i},$$

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:

$$\begin{bmatrix} \widehat{\alpha_i} \\ \widehat{\beta_i} \end{bmatrix}_{t+1} = \begin{bmatrix} \widehat{\alpha_i} \\ \widehat{\beta_i} \end{bmatrix}_{t} - \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}_{t}^{-1} \times \begin{bmatrix} L_{11} \\ L_{21} \end{bmatrix}_{t}. \qquad (4)$$

5.2.1.1 *Discussion.* Algorithm 1 shows the overall process for computing $\xi_i = (\alpha_i, \beta_i)$ for all items $i \in \{1, \ldots, n\}$. The overall process starts with the partitioning of the set of N users into K groups based on users' attitude. The partitioning is done using the `PartitionUsers` routine. This routine implements a one-dimensional clustering, 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 $\mathcal{O}(nN)$. Given these values, the Newton-Raphson estimation is performed for each one of the n items. This takes $\mathcal{O}(nIK)$ time in total, where I is the number of iterations for the estimation of one item. Therefore, the total running time of the 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 would suggest.

5.2.2 *The EM Algorithm for Item-Parameter Estimation.* In the previous section, we showed how to estimate item parameters $\xi_i = (\alpha_i, \beta_i)$ assuming that user attitudes $\vec{\theta} = (\theta_1, \ldots, \theta_N)$ were part of the input. In this section, we show how we can do the same computation without knowing $\vec{\theta}$. In this case, the input only consists of the response matrix $\mathbf{R}$. If $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ is the vector of parameters for all items, then our goal is to estimate $\vec{\xi}$ given response matrix $\mathbf{R}$, with $\vec{\theta}$ being the hidden and unobserved variables. We perform this task using the Expectation-Maximization (EM) procedure.

The EM procedure is an iterative method that consists of two steps: *expectation* and *maximization*. We describe these two steps below.

Algorithm 1. Item-parameter estimation of $\xi_i = (\alpha_i, \beta_i)$ for all items $i \in \{1, \ldots, n\}$.

```
Input: Response matrix $\mathbf{R}$, users attitudes $\vec{\theta} = (\theta_1, \ldots, \theta_N)$ and the number $K$ of users' attitude groups.
Output: Item parameters $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and $\vec{\beta} = (\beta_1, \ldots, \beta_n)$.
$\{F_g, \theta_g\}_{g=1}^{K} \leftarrow$ PartitionUsers($\vec{\theta}$, $K$)
for $g = 1$ to $K$ do
    $f_g \leftarrow |F_g|$
    for $i = 1$ to $n$ do
        $r_{ig} \leftarrow |\{j \mid j \in F_g$ and $\mathbf{R}(i, j) = 1\}|$
for $i = 1$ to $n$ do
    $(\alpha_i, \beta_i) \leftarrow$NR_Item_Estimation$\left(\mathbf{R}_i, \{f_g, r_{ig}, \theta_g\}_{g=1}^{K}\right)$
```

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

$$\mathbf{E}\left[f_g\right] = \overline{f}_g = \sum_{j=1}^{N} \mathbf{P}\left(\theta_g \mid \mathbf{R}^j, \vec{\xi}\right) \tag{5}$$

and

$$\mathbf{E}\left[r_{ig}\right] = \bar{r}_{ig} = \sum_{j=1}^{N} \mathbf{P}\left(\theta_g \mid \mathbf{R}^j, \vec{\xi}\right) \times \mathbf{R}(i, j)\,. \tag{6}$$

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.

5.2.2.2 *Maximization Step.* Knowing the values of $\overline{f}_g$ and $\bar{r}_{ig}$ 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 5.2.1.

The pseudocode for the EM algorithm is given in Algorithm 2. Every iteration of the algorithm consists of an *Expectation* and a *Maximization* step.

5.2.2.3 *The Posterior Probability of Attitudes.* By the definition of probability, this posterior probability is

$$\mathbf{P}\left(\theta_j \mid \mathbf{R}^j, \vec{\xi}\right) = \frac{\mathbf{P}\left(\mathbf{R}^j \mid \theta_j, \vec{\xi}\right) \mathbf{g}(\theta_j)}{\int \mathbf{P}\left(\mathbf{R}^j \mid \theta_j, \vec{\xi}\right) \mathbf{g}(\theta_j)\, d\theta_j}. \tag{7}$$

Function $\mathbf{g}(\theta_j)$ 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 [Mislevy and Bock 1986], we assume that the prior distribution $\mathbf{g}(.)$ is

Algorithm 2. 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 (6).
    // 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
```

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 $\mathbf{g}(X_t)$, the density of the attitude function at attitude value X_t. Then, we let $A(X_t)$ be the *area* of the rectangle defined by the points $(X_t - 0.5, 0)$, $(X_t + 0.5, 0)$, $(X_t - 0.5, \mathbf{g}(X_t))$, and $(X_t + 0.5, \mathbf{g}(X_t))$. Then the $A(X_t)$ values are normalized such that $\sum_{t=1}^{K} A(X_t) = 1$. In that way, we can obtain the posterior probabilities of X_t:

$$\mathbf{P}\left(X_t \mid \mathbf{R}^j, \vec{\xi}\right) = \frac{\mathbf{P}\left(\mathbf{R}^j \mid X_t, \vec{\xi}\right) A(X_t)}{\sum_{t=1}^{K} \mathbf{P}\left(\mathbf{R}^j \mid X_t, \vec{\xi}\right) A(X_t)}. \tag{8}$$

5.2.2.4 *Discussion.* The estimation of the privacy score using the IRT model requires as input the number of groups of users K. In our implementation, we follow standard conventions [Mislevy and Bock 1986] 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 $\mathcal{O}(Nn)$ 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 5.2.1 we know that $T_{\text{Max}} = \mathcal{O}(nIK)$, where I is the number of iterations of Equation (4). Again, Steps 12, 13, and 14 can be done in parallel due to the independence assumption of items.

5.3 IRT-Based Computation of Visibility

The computation of visibility requires the evaluation of $P_{ij} = \text{Prob}\left\{\mathbf{R}(i,j) = 1\right\}$, 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 of individuals, given 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 2.). 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

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

Since $\vec{\alpha}$ and $\vec{\beta}$ are part of the input, the only variable to maximize over is θ_j. The estimate of θ_j, denoted by $\widehat{\theta}_j$, is obtained iteratively using again the Newton-Raphson method. More specifically, the estimate $\widehat{\theta}_j$ at iteration $(t+1)$, $\left[\widehat{\theta}_j\right]_{t+1}$, is computed using the estimate at iteration t, $\left[\widehat{\theta}_j\right]_t$, as follows:

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

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

5.4 Putting It All Together

The sensitivity β_i computed in Section 5.2 and the visibility P_{ij} computed in Section 5.3 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, that is, sensitivity, attitude, and visibility, have

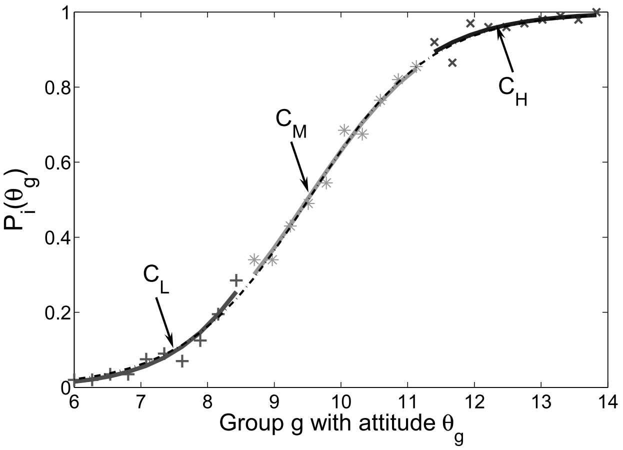


Fig. 2. True item characteristic curves (ICCs) and estimated ICCs using three different groups of users. y axis: $P_{ij} = P_i\,(\theta_j)$; x axis: ability level θ_j.

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 the 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 sensitivity of the same profile item estimated from different social networks is close to the same true value, and consequently, the privacy scores computed across different social networks are comparable.

The following example illustrates this group invariance property of IRT.

Example 3. In Figure 2 the dotted line is the true ICC for item i with (known) parameters $\xi_i = (\alpha_i, \beta_i)$. The data plotted by the markers in this figure consists of 30 groups; each marker depicts the proportion of people in a group that disclose item i. Since ICC spans the whole spectrum of attitude values, in order to estimate ξ_i from the data we need to consider all 30 groups. Instead of that, we estimate three pairs of item parameters ξ_i^L, ξ_i^M, and ξ_i^H using three nonoverlapping clusters of users with low, medium, and high attitudes, respectively; each cluster consists of 10 groups. We thus obtain three different ICCs: C_L, C_M, and C_H. In the figure we only draw (with solid lines) the fragments of these curves projected on the users that were actually used for estimating the corresponding item parameters. The *group invariance* property says that, as long as the three estimations consider responses to the same item, the fragments of C_L, C_M, and C_H should belong to the same (in this case the true) curve. Therefore, the estimated parameters ξ_i^L, ξ_i^M, and ξ_i^H should all be very close to the true value ξ_i. This is exactly what happens in Figure 2.

6. 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 $\mathbf{R}(i, j) = k$ with $k \in \{0, 1, \ldots, \ell\}$. The smaller the value of $\mathbf{R}(i, j)$, 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 be generalized as follows.

Definition 3. 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 4. The visibility of item i that belongs to user j at level k is denoted by $\mathrm{V}(i, j, k)$. The observed visibility is computed as $\mathrm{V}(i, j, k) = \mathbf{I}_{(\mathbf{R}(i,j)=k)} \times k$. The true visibility is computed as $\mathrm{V}(i, j, k) = P_{ijk} \times k$, where $P_{ijk} = \mathrm{Prob}\{\mathbf{R}(i, j) = k\}$.

Given Definitions 3 and 4, we compute the privacy score of user j using the following generalization of Equation (1):

$$\mathrm{P}\textsc{r}(j) = \sum_{i=1}^{n} \sum_{k=0}^{\ell} \beta_{ik} \times \mathrm{V}(i, j, k). \tag{9}$$

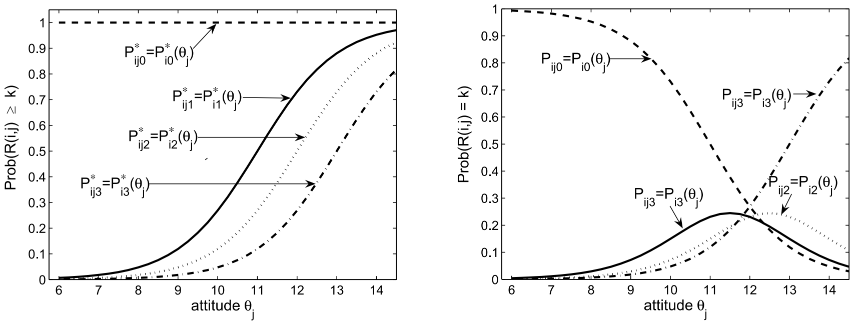
Again, in order to keep our framework more general, in the following sections, we will discuss true rather than observed visibility for the polytomous case.

6.1 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 $\mathbf{R}_k^*(i, j) = 1$, that is, $P_{ijk}^* = \mathrm{Prob}\{\mathbf{R}_k^*(i, j) = 1\} = \mathrm{Prob}\{\mathbf{R}(i, j) \geq k\}$. When $k = 0$, matrix $\mathbf{R}_{ik}^*$ has all its entries equal to 1, 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,

$$P_{ijk}^* = \frac{1}{1 + e^{-\alpha_{ik}^*(\theta_j - \beta_{ik}^*)}}. \tag{10}$$

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



(a) $P^*_{ijk} = P^*_{ik}(\theta_j)$ (b) $P_{ijk} = P_{ik}(\theta_j)$

Fig. 3. (a): y axis: probability $P^*_{ijk} = \text{Prob}\{\mathbf{R}(i, j) \geq k\}$ for $k \in \{0, 1, 2, 3\}$; x axis: attitude θ_j of user j. (b): y axis: probability $P_{ijk} = \text{Prob}\{\mathbf{R}(i, j) = k\}$ for $k \in \{0, 1, 2, 3\}$; x axis: attitude θ_j of user j.

levels $k \in \{1, \ldots, \ell\}$ do not cross, as shown in Figure 3(a). This observation results in the following corollary.

COROLLARY 1. *For items i and privacy levels* $k \in \{1, \ldots, \ell\}$, *we have that* $\beta^*_{i1} < \cdots < \beta^*_{ik} < \cdots < \beta^*_{i\ell}$. *Moreover, since curves* P^*_{ijk} *do not cross, we also have that* $\alpha^*_{i1} = \cdots = \alpha^*_{ik} = \cdots = \alpha^*_{i\ell} = \alpha^*_i$. *For* $k = 0$, $P^*_{ijk} = 1$, α^*_{i0} *and* β^*_{i0} *are not defined.*

The computation of the privacy score in the polytomous case, however, requires computing β_{ik} and $P_{ijk} = \text{Prob}\{\mathbf{R}(i, j) = k\}$ (see Definition 4 and Equation (9)). 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

$$P_{ijk} = \begin{cases} P^*_{ijk} - P^*_{ij(k+1)}, & \text{when } k \in \{0, \ldots, \ell - 1\}; \\ P^*_{ijk}, & \text{when } k = \ell. \end{cases} \tag{11}$$

Figure 3(b) shows the ICCs for P_{ijk}, which are obtained by the above equation.

Also, by Baker and Kim [2004], we also have the following proposition for β_{ik}.

PROPOSITION 1 BAKER AND KIM [2004]. *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* $\beta_{i0} < \beta_{i1} < \ldots < \beta_{i\ell}$.

Corollary 2 verifies our intuition that the sensitivity of an item is a monotonically increasing function of the privacy level k.

6.1.1 *IRT-Based Sensitivity for Polytomous Settings.* The sensitivity of item i with respect to privacy level k, β_{ik}, is the sensitivity parameter of the

P_{ijk} curve. We compute it by first computing the sensitivity parameters β^*_{ik} and $\beta^*_{i(k+1)}$. Then we use Proposition 1 to compute β_{ik}.

The goal here is to compute the ℓ sensitivity parameters $\beta^*_{i1}, \ldots, \beta^*_{i\ell}$ for each item i. As in Section 5, we consider two cases: one where the users' attitudes $\vec{\theta}$ are given as part of the input along with the response matrix $\mathbf{R}$, and the case where the input consists only of $\mathbf{R}$. We devote the rest of this section to discussing the algorithm for the first case. The second case can be solved using the same EM principles described in Section 5.2.2.

Given attitude vector $\vec{\theta} = (\theta_1, \ldots, \theta_N)$ as the input, one could argue that for every $k = 1, \ldots, \ell$ one could use Algorithm 1. with input $\mathbf{R}^*_k$ to compute the item parameters $\left(\alpha^*_{ik}, \beta^*_{ik}\right)$ for each level k and item i. Such a solution would give the wrong results for the following reasons: first, for each value of k, a different value for the discrimination parameter α^*_{ik} would be found. Second, the dependency of the P^*_{ijk} functions would not be taken into consideration.

These problems can be eliminated by simultaneously computing all $(\ell + 1)$ unknown parameters α^*_i and β^*_{ik} for $1 \leq k \leq \ell$. Again assume that the set of N individuals can be partitioned into K *groups*, such that all the individuals in the gth group have the same attitude θ_g. Also let $P_{ik}\left(\theta_g\right)$ be the probability that an individual j in group g sets $\mathbf{R}(i, j) = k$. Finally, denote by f_g the total number of users in the gth group and by r_{gk} the number of people in gth group that set $\mathbf{R}(i, j) = k$. Given this grouping, the likelihood of the data in the polytomous case can be written as

$$\prod_{g=1}^{K} \frac{f_j!}{r_{g1}!r_{g2}!\ldots r_{g\ell}!} \prod_{k=1}^{\ell} \left[P_{ik}\left(\theta_g\right)\right]^{r_{gk}}.$$

After ignoring the constants, the corresponding log-likelihood function is

$$L = \sum_{g=1}^{K} \sum_{k=1}^{\ell} r_{gk} \log P_{ik}\left(\theta_g\right). \tag{12}$$

To evaluate Equation (12), we use Equations (11) and (10). This substitution transforms L to a function where the only unknowns are the $(\ell + 1)$ parameters $(\alpha^*_i, \beta^*_{i1}, \ldots, \beta^*_{i\ell})$. The computation of these parameters is done using again an iterative Newton-Raphson procedure. The algorithm is similar to the one described in Section 5.2.1. The difference here is that there are more unknown parameters with respect to which we need to compute the partial derivatives of log-likelihood L given in Equation (12). Details can also be found in Baker and Kim [2004], Chapter 8.

6.1.2 *IRT-Based Visibility for Polytomous Settings.*

Computing the visibility values in the polytomous case requires the computation of the attitudes $\vec{\theta}$ for all individuals. Given the item parameters $\alpha^*_i, \beta^*_{i1}, \ldots, \beta^*_{i\ell}$, this can be done independently for each user, using a procedure similar to `NR_Attitude_Estimation` (see Section 5.3). The only difference here is that the likelihood function used for the computation is the one given in Equation (12).

6.1.3 *Putting It All Together.* The IRT-based computations of sensitivity and visibility for polytomous response matrices give a privacy score for every user. This score is computed by applying the IRT-based sensitivity and visibility values to Equation (9). As in the dichotomous IRT computations, 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.

7. 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.

7.1 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,

$$\beta_i = \frac{N - |\mathbf{R}_i|}{N}. \tag{13}$$

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

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

$$\beta_{ik}^* = \frac{N - \sum_{j=1}^{N} \mathbf{I}_{(\mathbf{R}(i,j) \geq k)}}{N}. \tag{14}$$

In order to be symmetric with the IRT-based computations for the polytomous settings, we compute the sensitivity value associated with level k, β_{ik}, by combining the β_{ik}^* and the $\beta_{i(k+1)}^*$ values as in Proposition 1. Note that, in this way, we guarantee that, for $k \in \{0, \ldots, \ell\}$, $\beta_{i0} < \beta_{i1} < \ldots < \beta_{i\ell}$ as required by Definition 3.

7.2 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 $|\mathbf{R}^j|$ is the number of items for which j sets $\mathbf{R}(i, j) = 1$, we have

$$P_{ij} = \frac{|\mathbf{R}_i|}{N} \times \frac{|\mathbf{R}^j|}{n}. \tag{15}$$

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:

$$P_{ijk} = \frac{\sum_{j=1}^{N} \mathbf{I}_{(\mathbf{R}(i,j)=k)}}{N} \times \frac{\sum_{i=1}^{n} \mathbf{I}_{(\mathbf{R}(i,j)=k)}}{n}. \tag{16}$$

The Naive computation of privacy score requires applying Equations (13) and (15) to Equation (1). For the polytomous case, we use Equation (9) to combine the β_{ik} and the P_{ijk} values computed as described above. We refer to the privacy score computed in this way as the **Pr_Naive** *score*.

7.3 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 **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. Therefore, the Naive approach does not exhibit the nice *group invariance* property. Moreover, as we will show in the experimental section, the probability model defined by Equations (15) and (16), though simple and intuitive, fails to fit the real-world response matrices **R** (in terms of χ^2 goodness-of-fit).

8. NETWORK-BASED PRIVACY SCORE

So far, we have defined the visibility of an item so that it only depends on the privacy setting picked by a user. In fact, the visibility of a profile item also depends on the position of a user within the social network. That is, if a popular user makes one of the items in his profile accessible by everyone, then this item becomes much more visible or her compared to the corresponding item of a more isolated user that is also publicly available.

Formally, so far we assumed that for user j the *visibility* of item i at level k is simply the product $\text{Prob}\{\mathbf{R}(i,j) = k\} \times k$, with $k \in \{0, \ldots, \ell\}$. We can generalize this definition to

$$\text{V}(i,j,k) = \text{Prob}\{\mathbf{R}(i,j) = k\} \times f_j(k), \tag{17}$$

where $f_j()$ is a monotonically increasing function of k with $f_j(0) = 0$. Note that the form of $f_j()$ depends on user j. For example, given a social network $\mathcal{G}$, $f_j()$ can depend on the position of j in $\mathcal{G}$. Equation (17) assumes that function $f_j(k)$ takes the same value for all items i. A more general setting would be one where this function is different not only for every user j, but also for every item i. For simplicity of exposition, we assume the former scenario.

Given $\mathcal{G}$, we evaluate $f_j(k)$ by exploiting notions from information-propagation models used in social-network analysis [Kempe et al. 2003]. In this setting, $f_j(k)$ should be interpreted as the fraction of nodes in the network that know the value of item i for user j, given that $\mathbf{R}(i, j) = k$. For $\mathbf{R}(i, j) = 0$, $f_j(0) = 0$; naturally a piece of information that is not released cannot be spread in the network. For $k = 1$, information about item i propagates from j to j's friends in $\mathcal{G}$, and from them to other users of $\mathcal{G}$. Let $\mathcal{P}$ be a *propagation model* that determines how information propagates from one node to its neighbors in $\mathcal{G}$. Also let $\mathcal{P}(j, \mathcal{G})$ be the *fraction of nodes in* $\mathcal{G}$ that know a piece of information about j once j releases it.[7] We define $f_j(1)$ to be $\mathcal{P}(j, \mathcal{G})$. In order to compute $f_j(k)$ for $k \geq 2$, we extend the original graph $\mathcal{G}$ to $\mathcal{G}^k$ by adding directed links from j to all the nodes in $\mathcal{G}$ that are within distance k from j. We then perform propagation of information from j to the graph $\mathcal{G}^k$ and let $f_j(k) = \mathcal{P}\left(j, \mathcal{G}^k\right)$.

In our experiments, we set the propagation model $\mathcal{P}$ to be the *Independent Cascade* (IC) model (see Kempe et al. [2003] for a more thorough description of the model). In this model, propagation proceeds in discrete steps. When node v gets a piece of information for the first time at time t, it is given a single chance to pass the information to each one of its currently oblivious immediate neighbors. Node v succeeds in passing the information to node w with probability $p_{v,w}$. If v succeeds, w gets to know the piece of information at time $t + 1$. Independently of whether v succeeds, it cannot make any further attempts to pass the information to w in subsequent rounds. For our experiments, we assumed that $p_{v,w}$ is the same for all neighboring nodes. Alternatively, one can use the information about the attitude of users and the IRT model to determine these probabilities. From the implementation point of view, one can compute $f_j(k) = \mathcal{P}\left(j, \mathcal{G}^k\right)$ by sampling every edge $(v \rightarrow w)$ of graph $\mathcal{G}^k$ with probability $p_{v,w}$. Implementation details on how to compute $f_j(k)$ for IC can be found in Kempe et al. [2003].

Having computed $f_j(k)$ using the IC model, we can then compute visibility $\mathrm{V}(i, j, k)$ using Equation (17). Combining this visibility with the appropriate sensitivity values, we can estimate the privacy score of users using Equation (9). When $\text{Prob}\left\{\mathbf{R}(i, j) = k\right\}$ and β_{ik} are computed using the Naive model described in Section 7, then we refer to the obtained score as the **Pr_Naive_IC** *privacy score*. When the computation of $\text{Prob}\left\{\mathbf{R}(i, j) = k\right\}$ and β_{ik} is done using the IRT model described in Section 6.1, we refer to the obtained score as the **Pr_IRT_IC**.

Note that our model is not restricted to the information-propagation models described above. In fact, any other of the information-propagation models described in Kempe et al. [2003] could be used to compute the visibility of a node as well.

9. EXPERIMENTS

The purpose of the experimental section is to illustrate the properties of the different methods for computing users' privacy scores and pinpoint their

[7] $\mathcal{P}(j, \mathcal{G})$ can either refer to the *actual* or the *expected* fraction depending on whether the propagation model $\mathcal{P}$ is *deterministic* or *probabilistic*, respectively.

advantages and disadvantages. From the data analysis point-of-view, our experiments with real data show interesting facts about users' behavior.

9.1 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 $\mathbf{R}_S$ was generated as follows: for each item i, of a total of $n = 30$ items, we picked parameters α_i and β_i uniformly at random from intervals $(0, 2)$ and $[6, 14]$, respectively. We assumed that the items were sorted based on their β_i values, that is, $\beta_1 < \beta_2 < \ldots < \beta_n$. Next, $K = 30$ different attitude values were picked uniformly at random from the real interval $[6, 14]$. Each such attitude value θ_g was associated with a group of 200 users (all 200 users in a group had 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 conducting an online survey. The goal of the survey was 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 were asked to specify the extent to which they wanted to share each item with others. The privacy levels a user could allocate to items were $\{0, 1, 2, 3, 4\}$; **0** means that a user wanted to share this item with no one, **1** with some immediate friends, **2** with all 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 collected information about their locations, educational backgrounds, ages, 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 range of 23 to 39, 91.6% hold a college degree or higher, and 76.0% spend 4 h or more everyday surfing online.

From the Survey dataset we constructed 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 constructed 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.

We conducted the survey on SurveyMonkey[8] for 3 months in order to obtain the users' answers to the questions. However, due to privacy concerns and IBM's policy, we are not currently allowed to make the dataset publicly.

[8] http://www.surveymonkey.com/

9.2 Experiments with Dichotomous Synthetic Data

The goal of the experiments described in this section was to demonstrate the group invariance property of the IRT model. For the experiments, we used the Dichotomous Synthetic dataset.

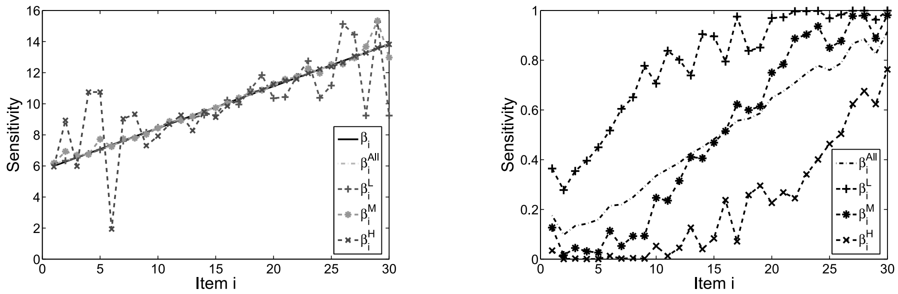
We conducted the experiments as follows: first, we clustered the 6000 users into three groups $F_L = \cup_{g=1\ldots10} F_g$, $F_M = \cup_{g=11\ldots20} F_g$, and $F_H = \cup_{g=21\ldots30} F_g$. That is, the first cluster consists of users in the 10 lowest-attitude groups $F_1, \ldots, F_{10}$, the second consists of all users in the 10 medium-attitude groups, and the third consists of all users in the 10 highest-attitude groups. Given users' attitudes assigned in the data generation, we estimated 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 was done using Algorithm 1 with an input response matrix that only contained the columns of $\mathbf{R}_S$ associated with the users in F_L, F_M, and F_H respectively. We also used Algorithm 1 to compute estimates $\xi_i^{all} = \left(\alpha_i^{all}, \beta_i^{all}\right)$ using the whole response matrix $\mathbf{R}_S$.

Figure 4(a) shows the estimated sensitivity values of the items. Since the data was generated using the IRT model, the true parameters $\xi_i = (\alpha_i, \beta_i)$ for each item were 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 show what is known as group invariance. The validity of this property was demonstrated in Frank Baker's book [Baker and Kim 2004] and in an online tutorial.[9] 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 and 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 4(a) there are some items for which the estimated parameters deviate from the true one more.

We repeated the same experiment for the Naive model. That is, for each item we estimated sensitivities β_i^L, β_i^M, β_i^H, and β_i^{all} using the Naive approach (Section 7). Figure 4(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) were much higher than the β_i^{all} estimates since these users rarely shared anything, whereas β_i^H obtained

[9] `http://echo.edres.org:8080/irt/baker/chapter3.pdf`



(a) IRT model

(b) Naive model

Fig. 4. Testing the group-invariance property of item parameter estimation using the IRT (Figure 4(a)) and Naive (Figure 4(b)) models.

from users with high-attitude levels (i.e., careless, extrovert) were much lower than β_i^{all}.

Note that since sensitivities estimated by the Naive and IRT models are not on the same scale, one should consider the relative error instead of absolute error when comparing the results in Figure 4(a) and 4(b).

9.3 Experiments with the Survey Data

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

9.3.1 *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 the 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 consisted of dichotomous matrix $\mathbf{R}_k^*$ ($k \in \{1, 2, 3, 4\}$) constructed from the Survey data.

First we tested whether the IRT model is a good model for data in $\mathbf{R}_k^*$. We tested this hypothesis as follows: first we used the EM algorithm (Algorithm 2) to estimate both items' parameters and users' attitudes. Then, we used a one-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. Also let the size of F_g be f_g. Next, for each item i and group g we computed

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

In this equation, f_g is the number of users in group F_g; p_{ig} (respectively, $\tilde{p}_{ig}$) is the expected (respectively 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

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

	$\mathbf{R}_1^*$	$\mathbf{R}_2^*$	$\mathbf{R}_3^*$	$\mathbf{R}_4^*$	
	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

$p_{ig} = P_i\left(\theta_g\right)$ and it is computed using Equation (2) for group attitude θ_g and item parameters estimated by EM. For IRT, the test statistic followed, approximately, a χ^2 distribution with $(K-2)$ degrees of freedom since there were two estimated parameters.

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

$$p_{ig} = \left(\frac{|\mathbf{R}_{k_i}^*|}{N}\right) \times \left[\frac{1}{f_g}\sum_{j\in F_g}\frac{|\mathbf{R}_k^{*^j}|}{n}\right], \tag{18}$$

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 followed 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 followed the IRT or Naive model was rejected. We show results for all 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 items followed the Naive model were rejected for all 49 items. On the other hand, the null hypothesis that items followed the IRT model was rejected for only a small number of items in all configurations. This indicates that the IRT model better fits the real data. All results reported here are for confidence level .95.

9.3.2 *Sensitivity of Profile Items.* In Figure 5 we visualize, using a tag cloud, the sensitivity of the profile items used in our survey. The evaluation of sensitivity values was done using the EM algorithm (Algorithm 2.) 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 was the most sensitive item, while Gender, which locates just right above the letter "h" of "Mother" has the lowest sensitivity, too small to be visually identified.

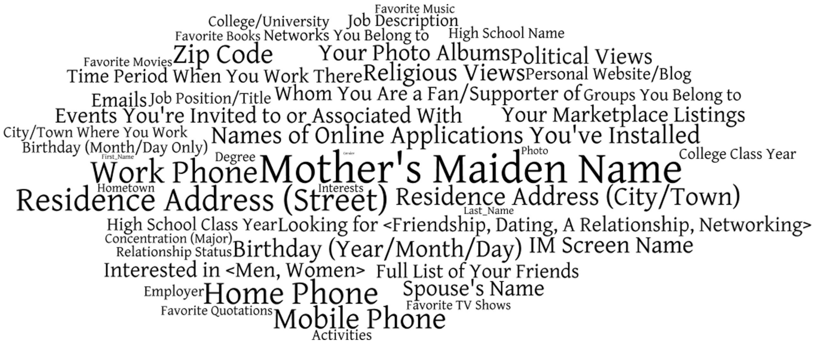


Fig. 5. Sensitivity of the profile items computed using IRT model with input the dichotomous matrix $\mathbf{R}_2^*$. Larger fonts mean higher sensitivity.

9.4 Comparison of Privacy Scores

The goal of this experiment was compare the privacy scores obtained using different scoring schemes. Since scores obtained using different methods were not on the same scale, we compared them using the Pearson correlation coefficient. We showed that IRT model produced more robust privacy scores than the Naive approach.

For this experiment we used the Survey dataset. Using as inputs the polytomous response matrix **R** and methods from Section 6, we obtained privacy scores **Pr_Naive** and **Pr_IRT**. Also, using as inputs the dichotomous matrix $\mathbf{R}_2^*$ and the Naive and IRT methods, we obtained scores **Pr_Naive*** and **Pr_IRT***, respectively.

We also computed privacy scores by taking into account information about the structure of the the users' social networks. We did so using the methodology described in Section 8. Unfortunately, the Survey data consists of responses of individuals to a set of survey questions and we are not aware of the underlying social-network structure. However, since we wanted to compare the privacy scores obtained using all the proposed scoring schemes, we constructed an artificial social network G among the respondents of our survey. The network G was constructed as follows: first we constructed five clusters of users, based on respondents' geographic location. These five clusters corresponded to users in North America—West coast, North America—East coast, Europe, Asia, and Australia. Each one of these clusters consisted of 71, 30, 29, 12, and 11 users, respectively. We added connections between respondents in the same clusters so as to generate a powerlaw graph between them. For this we used the graph-generation model described in Barabási and Albert [1999]. Finally, we connected the powerlaw subgraphs that correspond to each cluster by adding random links between nodes in different subgraphs with probability $p = 0.01$. Using graph G, response matrix **R** (respectively $\mathbf{R}_2^*$), and the methods from Section 8, we computed privacy scores **Pr_Naive_IC** and **Pr_IRT_IC** (respectively **Pr_Naive*****_IC** and **Pr_IRT*****_IC**).

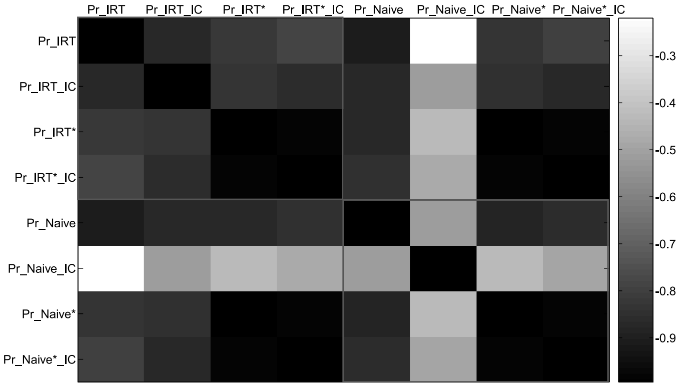


Fig. 6. Survey data: comparison of privacy scores using correlation coefficient; darker colors correspond to higher values of the correlation coefficient.

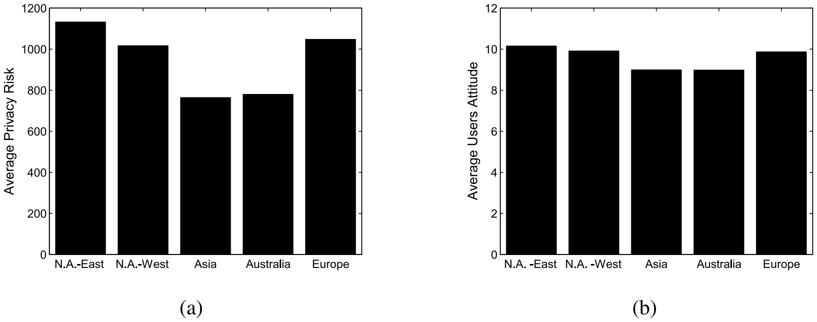


Fig. 7. Survey data: average privacy scores (**Pr_IRT**) (a) and average users' attitudes (b) per geographic region.

Figure 6 shows the values of the Pearson correlation between the privacy scores obtained using the eight aforementioned scores—darker colors correspond to higher correlation coefficients. Note that the 4×4 submatrix in the top left, which contains the correlation coefficients between the privacy scores computed using IRT, has consistently high-correlation values. Thus, the IRT model produced more robust privacy scores. On the other hand, the Naive model was not as consistent. For example, **Pr_Naive_IC** scores seem to be significantly different from all the rest.

9.4.1 *Geographic Distribution of Privacy Scores.* Here we present some interesting findings we got by further analyzing the Survey dataset. We computed the the privacy scores of the 153 respondents using the polytomous IRT-based computations (Section 6.1).

After evaluating the privacy scores of individuals using as input the whole response matrix **R**, we grouped the respondents based on their geographic locations. Figure 7(a) shows the average values of the users' **Pr_IRT** scores per location. The results indicate that people from North America and Europe had higher privacy scores (high risk) than people from Asia and Australia. Figure 7(b) shows the average users' attitudes per geographic region. The privacy scores and the attitude values are highly correlated. This experimental finding indicates that people from North America and Europe are more comfortable in revealing personal information on the social networks in which 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.

10. CONCLUSIONS

We have presented models and algorithms for computing the privacy scores of users in online social networks. Our methods take into account the privacy settings of users with respect to their profile items as well as their positions in the social network. Our framework uses notions from item response theory and information-propagation models. We described the mathematical underpinnings of our methods and presented a set of experiments on synthetic and real data that highlight the properties of our models and the current trends in users' behavior. We believe that our framework tackles the issue of privacy in online social networking from a new user-centered perspective and can prove useful in growing users' awareness.

REFERENCES

AHMAD, O. 2006. Privacy management method and apparatus. Patent application U.S. 2006/0047605.

BACKSTROM, L., DWORK, C., AND KLEINBERG, J. M. 2007. Wherefore art thou R3579X? Anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th International Conference on the World Wide Web (WWW)*. 181–190.

BAKER, F. B. AND KIM, S.-H. 2004. *Item Response Theory: Parameter Estimation Techniques*. Marcel Dekker, New York, NY.

BARABÁSI, A.-L. AND ALBERT, R. 1999. Emergence of scaling in random networks. *Science 286,* 5439, 509–512.

BIRNBAUM, A. 1968. Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores*, F. Lord and M. Novicle, Eds. Addison-Wesley, Reading, MA, 397–479.

FANG, L. AND LEFEVRE, K. 2010. Privacy wizards for social media sites. In *Proceedings of the International Conference on the World Wide Web (WWW)*.

GROSS, R. AND ACQUISTI, A. 2005. Information revelation and privacy in online social networks. In *Proceedings of the ACM Workshop on Privacy in the Electronic Society*. 71–80.

HAY, M., MIKLAU, G., JENSEN, D., TOWSLEY, D., AND WEIS, P. 2008. Resisting structural reidentification in anonymized social networks. *Proc. VLDB Endow. 1,* 1, 102–114.

KEMPE, D., KLEINBERG, J. M., AND TARDOS, É. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 137–146.

LEONARD, A. 2004. You are who you know. http://dir.salon.com/tech/feature/2004/06/15/social_software_one/index.html.

LIU, H. AND MAES, P. 2005. Interestmap: Harvesting social network profiles for recommendations. In *Proceedings of the Beyond Personalization Workshop*.

Liu, K. and Terzi, E. 2008. Towards identity anonymization on graphs. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*. 93–106.

Mislevy, R. and Bock, R. 1986. *PC-BILOG: Item Analysis and Test Scoring with Binary Logistic Models*. Scientific Software, Mooreville, IN.

Owyang, J. 2008. Social network stats: Facebook, myspace, reunion.
http://www.web-strategist.com/blog/2008/01/09/.

Richardson, M., Agrawal, R., and Domingos, P. 2003. Trust management for the Semantic Web. In *Proceedings of the International Semantic Web Conference*. 351–368.

Ying, X. and Wu, X. 2008. Randomizing social networks: A spectrum preserving approach. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*. 739–750.

Ypma, T. J. 1995. Historical development of the Newton-Raphson method. *SIAM Rev. 37,* 4, 531–551.

Zhou, B. and Pei, J. 2008. Preserving privacy in social networks against neighborhood attacks. In *Proceedings of the 24th International Conference on Data Engineering (ICDE)*. 506–515.

Received October 2009; revised March 2010; accepted April 2010