Latent-Data Privacy Preserving With Customized Data Utility for Social Network Data

Zaobo He, *Student Member, IEEE*, Zhipeng Cai, *Senior Member, IEEE*, and Jiguo Yu

***Abstract*—Social network data can help with obtaining valuable insight into social behaviors and revealing the underlying benefits. New big data technologies are emerging to make it easier to discover meaningful social information from market analysis to counterterrorism. Unfortunately, both diverse social datasets and big data technologies raise stringent privacy concerns. Adversaries can launch inference attacks to predict sensitive latent information, which is unwilling to be published by social users. Therefore, there is a tradeoff between data benefits and privacy concerns. In this paper, we investigate how to optimize the tradeoff between latent-data privacy and customized data utility. We propose a data sanitization strategy that does not greatly reduce the benefits brought by social network data, while sensitive latent information can still be protected. Even considering powerful adversaries with optimal inference attacks, the proposed data sanitization strategy can still preserve both data benefits and social structure, while guaranteeing optimal latent-data privacy. To the best of our knowledge, this is the first work that preserves both data benefits and social structure simultaneously and combats against powerful adversaries.**

***Index Terms*—Data sanitization, latent-data privacy, prediction utility loss, structure utility loss, tradeoff.**

I. Introduction

AMONG the many big data resources, social networks contribute considerable amount of data covering all the aspects of frontend and backend. Facebook has 1.65 billion users with 1 billion active users per month, Twitter has 600 million users with 0.5 billion tweets published per day, Amazon has 304 million users with 9.65 billion items traded per year, Tencent QQ has 829 million active users with up to 210 million simultaneous online users, WeChat has over a billion users with 700 million active users, etc. With such large scale of and variety of data, Social Network Analysis (SNA) becomes increasingly important for classifying end users, predicting buying interests, foretelling event occurrence, etc. Recent years have witnessed the boom of social networks, offering a great opportunity for SNA to prompt more novel applications.

Although the abundant social data bring valuable benefits, they unfortunately raise stringent privacy concerns as well. Each social network user is generally associated with an attribute set which may contain sensitive attributes like location, gender, sexual orientation, etc. Such personal information could be exploited by third parties like data analysts, marketer, or social media itself. Any third parties with malicious intentions on sensitive information of users can be viewed as adversaries and they breach user privacy by collecting sensitive data first. People now begin to concern about the privacy issue and become more conservative in publishing personal and sensitive data, which may degrade data publishing scale and drive users to publish anonymized data. Therefore, the conflict between privacy concerns and data utility promotes adversaries to exploit sensitive information contained in the published data.

Concerns derived from inference attacks towards sensitive information contained in user data is represented as *latent-data privacy*, where the inference attacks usually employ statistical analysis, machine learning or data mining techniques to infer sensitive information. For instance, suppose a user does not disclose her opinions and interests online. Unfortunately, it is easy to predict some of her opinions and interests if it is publicly known that she is affiliated with any particular organization or club. ABCNews.com and Boston Globe [1] shown it is achievable to infer the sexual orientation of a user through mining a Facebook subnetwork involving the user's friendship relations, gender, and other attributes. Latent-data privacy breaches could incur serious negative repercussions.

Publishing sanitized data is generally adopted to protect latent-data privacy. Data sanitization methods introduce noises by sanitizing attribute sets or social links. Although sanitizing publicly available data can help with protecting latent-data privacy, such simple methods could also reduce data utility for SNA. On the one hand, some user attributes are indicative for specific social analysis which is expected to be accurately predicted. For instance, a SNA server utilizes published Facebook data to make movie recommendation for target users. Unfortunately, some dominant attributes, such as "gender", may have been sanitized to protect latent-data privacy, degrading recommendation performance. On the other hand, in addition to sanitizing attributes, sanitizing social network links can distort friendship relations among users and change one's social status, which is another reason of reducing data utility for SNA. For example, social link sanitization can

Manuscript received February 28, 2017; revised June 27, 2017; accepted July 24, 2017. Date of publication August 11, 2017; date of current version January 15, 2018. This work was partly supported by the National Science Foundation (NSF) under Grants 1252292, 1741277, and 1704287, and the NSF of China under Contracts 61373083, 61370084, 61502116, 61371185, and 61373027, the NSF of Shandong Province under Contract ZR2012FM023, and China Postdoctoral Science Foundation NO.2015M571231. The review of this paper was coordinated by Prof. J. Weng. *(Corresponding author: Zhipeng Cai.)*

Z. He and Z. Cai are with the Department of Computer Science, Georgia State University, Atlanta, GA 30302 USA (e-mail: zhe4@student.gsu.edu; zcai@gsu.edu).

J. Yu is with the School of Information Science and Engineering, Qufu Normal University, Rizhao 276826, China (e-mail: jiguoyu@sina.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2017.2738018

turn an influential user to an unsocial one. Therefore, effective privacy preserving SNA strategies are crucial for big social network data.

In this work, we explore how to balance the tradeoff between latent-data privacy and data utility. We assume adversaries collect user data, and some privacy-unconscious users publish their sensitive latent information. We first formalize the metrics to measure data utility loss and latent-data privacy. Then, we propose two data sanitization methods that sanitize social attributes and links, respectively. Finally, data-sanitization strategies are proposed, which should not degrade the benefits brought by social network data, while sensitive latent information can still be protected.

To measure data utility loss, we introduce prediction accuracy deviation and network structure disparity. Both of them cause utility loss because of the employed data sanitization strategy. We investigate how to measure them and their relationship. Previous works usually consider them separately. Network structure disparity not only affects prediction accuracy, but also limits social interaction among users. The current metrics do not comprehensively measure data utility and could sacrifice more utility in realizing privacy-utility tradeoff. Our work does consider both prediction accuracy deviation and network structure disparity. For latent-data privacy, we expect our data sanitization strategy can combat against powerful adversaries with abundant prior knowledge who launch inference attacks. Thus, it is necessary to figure out how adversaries launch inference attacks. Previous works primarily assume relatively weak adversaries such that the proposed data sanitization strategy is not effective. Our work does consider this problem and quantify the capabilities of adversaries.

The previous studies for privacy-utility tradeoff have several deficiencies. First, attribute-sanitization and link-sanitization are separately considered, degrading the privacy preserving effect. Second, relatively weak adversaries are assumed so that the proposed data sanitization strategies are not sufficient to combat against powerful adversaries. Third, structure utility loss caused by social structure disparity is ignored so that preserved utility is overestimated. Therefore, the previous studies cannot effectively optimize the tradeoff between latent-data privacy and data utility. In this paper, we identify an optimization problem seeking a data sanitization strategy to realize the maximum latent-data privacy with customized data utility. Our main contributions are summarized as follows:

1) We consider prediction utility loss and structure utility loss simultaneously rather than separately.
2) We assume powerful adversaries who can launch optimal inference attacks instead of weak adversaries.
3) Rather than separately considering attribute-sanitization and link-sanitization, we collectively sanitize social links and attributes.

We organize the paper as follows. Section II addresses the related works. Section III introduces Network model and problem definition. Section IV introduces the prediction method for latent attributes and data-sanitization method. In Section V, privacy and utility metrics are introduced. The data sanitization strategy to optimize the privacy-utility tradeoff is presented in Section VI. The performance evaluation are shown in Section VII. Section VIII concludes the paper.

II. Related Works

Privacy threats towards social network data have been extensively documented. A large body of researches investigating the attacks against anonymous social network data, diverse in techniques or goals, have been performed. On the other hand, inference attacks on published social network data have attracted much attention. The threat of predicting sensitive information become now a serious issue due to the popularity of social networks.

Many previous works investigate how to predict sensitive latent information. In [2], several link-prediction and attribute-prediction algorithms are proposed in social-attribute networks. In [3], the authors employ the big data technologies to predict demographic information of users such as age and location based on users' mobile communication patterns. The work in [4] designs a method to predict sensitive latent information from texts published in social media. In [5], it is demonstrated that sensitive latent information could be predicted combining with community information with the observation that users with common attributes are more likely to be friends and often form dense communities. The work in [6] develops a data-sanitization strategy to predict sensitive information which can harness link and attribute information simultaneously. The work in [6] evaluates the effect of removing links, removing attributes and perturbing attributes on protecting sensitive latent information. Our previous work also [7] studies how to customize the tradeoff data utility and customized latent-data privacy in classification based applications.

Data sanitization is important for privacy protection to tune the privacy-utility tradeoff [8]–[26]. Both the work [8] and [9] sanitize data by synthesizing sampled data so that synthesized data satisfy differential privacy. In addition to sensitive latent information, protecting social network property privacy, like link privacy [10], degree distribution [11], graph privacy [12] and applications such as influence maximization [13] and privacy preserving content sharing [14], also attracts much attention. [15] explored how to sanitize data to optimize the tradeoff between three parties: utility, inherent-data privacy and latent-data privacy. To protect against inference attacks on social data, [16] proposed a data-sanitization method that can sanitize social attributes and links collectively with different sanitization methods. [17] explored the inference attacks on personal traits and genotypes based on belief propagation. Furthermore, a genomic data sanitization method is proposed in [17], by removing most indicative genomes to traits.

Existing privacy preserving techniques, like differential privacy [27], k-anonymity [28], l-diversity [29], are generally proposed for preserving inherent-data privacy; however, they are not competent for protecting latent-data privacy being subject to inference attacks. Inherent-data privacy is related to sensitive attribute contained in the attribute set released by users in order to receive data-related services. For example, age and gender are unavoidable data for health related services yet unwilling to be released by most consumers.

III. Problem Statement

A. Social Network Model

Definition 3.1. Social network: Social network is represented by graph model $G(V, E, \mathcal{X})$, with user set V, link set E, and the set of attribute sets, $\mathcal{X}$. For any link $e_{ij} \in E$ between users u_i and u_j, $e_{ij} \in E$ also indicates $e_{ji} \in E$.

Definition 3.2. Attribute set: For user $u_i \in V$, its attribute set is represented by an attribute vector $X_i \in \mathcal{X}$. Each attribute $x_j \in X_i$ $(1 \leq j \leq |X_i|)$ takes value(s) from the j-th dimension.

For social network data, a SNA server performs analysis to predict users' latent information such as preferences. Then, according to the predicted results, the corresponding services are provided. For example, a SNA server can predict movie preference of users by classifying the users into different classes such as *action*, *adventure*, *comedy*, etc. However, adversaries also attempt to gain benefit from users' social relationships and attribute set to infer sensitive latent information. These two types of latent information related to data utility and latent-data privacy are denoted as Sensitive Latent Attributes (SLA) and Non-Sensitive Latent Attributes (NSLA), respectively.

Definition 3.3. SLA: SLA is a set of unpublished sensitive attributes, yet such attributes could be predicted from published social network data combined with prior knowledge.

Definition 3.4. NSLA: NSLA is a set of unpublished non-sensitive attributes, yet such attributes can be predicted from published social network data combined with prior knowledge.

We expect NSLA can be accurately predicted so that satisfactory services can be guaranteed. Conversely, to protect the privacy of SLA, we expect SLA does not being predicted accurately. Furthermore, social network structure should be preserved such as node degree, centrality, betweenness, etc. Thus, there exists a tradeoff between latent-data privacy and data utility. Utility and latent-data privacy are formally defined as follows.

Definition 3.5. Latent-data privacy: Latent-data privacy preserving is to protect the SLA of each user.

Definition 3.6. Utility: The utility of a social network dataset is high iff 1) a SNA server has a high prediction accuracy for NSLA; and 2) the social network structure is effectively preserved.

For the sake of brevity, we omit the subscript and use X and X' to denote an original and sanitized attribute set of a user, respectively, in the rest of the paper without confusion.

B. Model of Adversaries

We assume powerful adversaries with abundant prior knowledge about users, and they can launch optimal inference attacks to infer the SLA of each user. This assumption allows the constructed data-sanitation method can combat against adversaries with a larger range of capability.

There exists a prior probability for a user's attribute vector X, denoted as $\psi(X)$, which represents the probability of a user with attribute set X. For a user, all her possible attribute sets satisfy $\sum \psi(X) = 1$. We call the set of $\psi(X)$ as a user's profile.

TABLE I
Major Symbols

Parameter	Definition
$\mathcal{X}$	Set of attribute sets
X_i	Attribute set of user u_i
x_j	j-th attribute
$\psi(X)$	Prior probability of attribute set X
l_t^i	t-th latent attribute, l_t, of u_i
$P(l_t^i)$	Probability of u_i with latent attribute l_t
$W_{i,j}$	Weight between u_i and u_j
$f(X'\|X)$	Attribute sanitization strategy
$\mathcal{L}(X'\|X)$	link sanitization strategy
ϵ	Structure-utility loss threshold
δ	Prediction-utility loss threshold

Definition 3.7. Profile: The profile of a user is a set of probabilities $\Psi = \{\psi(X_1), \psi(X_2), \ldots, \psi(X_k)\}$, $\sum_{1 \leq i \leq k} \psi(X_i) = 1$, where each $\psi(X_i)$ is the probability of a user with attribute set X_i and k is the number of possible attribute sets.

First, we assume adversaries know each user's profile. Second, adversaries are assumed to know the data-sanitization strategy employed to realize the tradeoff between utility and privacy. Based on the above knowledge, optimal inference attacks are launched by adversaries.

C. Problem Definition

In this paper, we study the following problem.

Input:

1) Social graph G, SLA and NSLA of users.
2) Utility thresholds ϵ and δ.

Output:

The data sanitization strategy that minimizes the predication accuracy for unpublished SLA and satisfies utility threshold ϵ and δ.

For clarity, the meanings of the symbols are summarized in Table I.

IV. Preliminaries

In this section, the prediction method is presented to predict both SLA and NSLA of a user based on published social data.

A. Prediction Method for Latent Attributes

We assume powerful adversaries that launch inference attacks by utilizing all publicly available knowledge including social links and attribute sets. Therefore, the prediction method predicts latent information considering social links and attribute sets collectively to increase prediction accuracy.

Link knowledge is important for predicting latent information in social networks. Therefore, we consider u_j' latent information when predicting u_i' latent information, where $u_j \in N_i$ and N_i denotes the neighbor set of u_i. For clarity, u_i with latent attribute l_t is denoted as l_t^i.

For brevity, the probability of u_i to have latent attribute l_t is denoted as $P(l_t^i)$. The average probability of u_i' neighbors with latent attribute l_t is calculated as:

$$P\left(l_t^i|N_i\right) = \frac{1}{|N_i|} \sum_{u_j \in N_i} P\left(l_t^j\right) \tag{1}$$

However, directly computing the average probability may incur overfitting. In practice, close neighbors should have larger impact for each other on the determination of latent information. To avoid overfitting, we introduce a weight to evaluate the impact of one neighbor for target user. We assume that if more published attributes are shared by two friends, they tend to share more latent attributes. Then the weight $W_{i,j}$ between u_i and u_j is calculated as

$$W_{i,j} = \frac{|\left(x_1^i, \ldots, x_m^i\right) \cap \left(x_1^j, \ldots, x_n^j\right)|}{|X_i|} \tag{2}$$

Equation (2) computes the proportion of the shared attributes between u_i and u_j among u_i's attributes. Clearly, $W_{i,j} \neq W_{j,i}$. To determine l^i based on N_i, we combine Equation (1) and Equation (2) as follows,

$$P\left(l_t^i|N_i\right) = \frac{1}{|N_i|} \sum_{u_j \in N_i} P\left(l_t^j\right) \frac{W_{i,j}}{\sum_{u_k \in N_i} W_{i,k}} \tag{3}$$

It is easy to find that Equation (3) requires that at least one of the neighbors of each user to publish her latent attributes. Obviously, this strict condition is hard to be satisfied in real social networks. Therefore, it is inaccurate to predict the latent attributes of user u_i based on link information directly, since it is possible that few neighbors publish their latent attributes. To solve this problem, we first predict the latent attributes of those unpublished users through analyzing their attribute sets. Then, we predict the latent attributes of unpublished users through utilizing weighted link knowledge calculated by Equation (3).

Next, we present how to predict the latent attributes of a user through analyzing her attribute set. Given a user u_i with attribute set $X_i = \{x_1, \ldots, x_n\}$ and p potential latent attributes $l_1, \ldots, l_p$, the probability of u_i with latent attribute l_t is $\arg\max\limits_{1 \leq t \leq p}[P(l_t^i|x_1, \ldots, x_n)]$.

To calculate the above value, based on Bayes Theorem, assuming that all attributes are independent, we have

$$\arg\max_{1 \leq t \leq p} \left[\frac{P\left(l_t^i\right) \times P\left(x_1|l_t^i\right) \times \ldots \times P\left(x_n|l_t^i\right)}{P(x_1, \ldots, x_n)}\right].$$

We find that $P(x_1, \ldots, x_n)$ is the same for any value of $P(l_t^i)$. Therefore, we only need to calculate

$$\arg\max_{1 \leq t \leq p} \left[P\left(l_t^i\right) \times P\left(x_1|l_t^i\right) \times \ldots \times P\left(x_n|l_t^i\right)\right].$$

B. Data Sanitization Method

In Section IV-A, we assume powerful adversaries that launch inference attacks by exploiting social links and attribute sets simultaneously. Therefore, in order to realize the tradeoff between privacy and utility, our objective is to sanitize both social links and attribute sets.

1) Attribute-Sanitization Method: An attribute set could be sanitized in three ways, *adding* attributes, *removing* attributes, and *perturbing* attributes (replace one attribute with another). Which methods should be employed to sanitize social data depends on data utility and privacy metrics and data semantics.

To prevent inference attacks on SLA, we can sanitize the most indicative attributes for each SLA which is publicly available to adversaries. With this objective, for a user with attribute set X, it is easy to determine the most indicative attribute x_j for any SLA $z_i \in Z$ by $\text{argmax}_j[\forall z_i \in Z : P(x_j|z_i)]$.

This allows us to determine a single attribute which is the most indicative for a SLA and sanitize it. Unfortunately, directly sanitizing the most indicative attributes for SLA can reduce utility if we don't consider the most indicative attributes for NSLA. For instance, consider the case to predict health conditions of users which could be viewed as NSLA. Health conditions and SLA such as sexual orientation share indicative attribute "gender". Therefore, although sanitizing "gender" reduces the prediction accuracy for SLA, it also reduces the prediction accuracy for NSLA.

To resolve the above conflict, we propose the following data sanitization method: (1) If there exist indicative attributes shared by SLA and NSLA, we *perturb* the shared indicative attributes; and *remove* the SLA except the shared indicative attributes; (2) If there does not exist any indicative attribute shared by SLA and NSLA, we *remove* the indicative attributes for SLA.

The next challenge is how to perturb the indicative attributes shared by SLA and NSLA. Our idea is to generalize each shared indicative attribute. For example, if a shared attribute is *idol: Jodon*, it can be generalized to *basketball star*. For each shared indicative attribute, we can organize potential generalized attributes into a hierarchy.

2) Link-Sanitization Method: Unlike attributes, social links can only be sanitized by *adding* links and *removing* existing links. Similar with the attribute-sanitization method, a link-sanitization method should reduce the prediction accuracy for SLA and do not greatly reduce the prediction accuracy for NSLA. Unfortunately, unlike attributes, it is nontrivial to find the indicative links shared by SLA and NSLA, thus we focus on reducing the prediction accuracy for SLA firstly when sanitize links and more constraints will be given later to guarantee utility.

For this goal, the concept of Vulnerable Link is introduced as follows:

Definition 4.1. Vulnerable link: A vulnerable link of one user is the link whose removal will lower the prediction accuracy for the SLA of the user. The prediction accuracy for the SLA of u_i upon removing the vulnerable link e_{ij} is $\Lambda(E_i - e_{ij})$.

From the above definition, it shows that $\Lambda(E_i - e_{ij}) \leq \Lambda(E_i)$. To protect SLA of u_i through removing links, we first identify a set of vulnerable links denoted as A_i. Second, for any $e_{ij} \in A_i$, we calculate the reduction of prediction accuracy for SLA upon removing e_{ij}. Then, we order the links in A_i according to the calculated prediction accuracy reduction. We next remove those links with the largest prediction accuracy reduction in A_i.

V. Metrics

Now we discuss how to measure utility and latent-data privacy. Our data sanitization strategy includes two parts: attribute sanitization strategy $f(X'|X)$ and link sanitization strategy $\mathcal{L}(X'|X)$. $f(X'|X)$ likes a transfer function that takes a user's attribute set X as input and outputs the sanitized one X'. Meanwhile, for an arbitrary user u_i, $\mathcal{L}(E_i'|E_i)$ can be viewed as a transfer function that takes u_i's link set E_i as input and outputs the sanitized one E_i'.

A. Utility Metric

For data utility, two aspects need to be considered. First, the sanitized attribute set and social links should guarantee a SNA server can effectively infer the NSLA of users. Second, the sanitized network structure should not deviate from the original one very much. Worth to note that the second aspect expects that sanitizing social links does not distort friendship relations among users and does not change one's social status too much. We introduce two parameters ϵ and δ to scale the above two aspects. Then, (ϵ, δ)-data utility can be defined as follows.

Definition 5.1. (ϵ, δ)-Utility: Given social graph G, network disparity measurer $\mathcal{M}$, collective prediction method $\mathcal{C}$, NSLA set Y, accessible prior knowledge known to third party users $\mathcal{K}$, we say that G's sanitized graph G' satisfies (ϵ, δ)-utility if for any NSLA $y_i \in Y$,

$$\begin{aligned}&(\mathrm{i}).\, \mathcal{M}(G, G') \leq \epsilon;\\ &(\mathrm{ii}).\, \Lambda_{\mathcal{C}}^{y_i}(G', \mathcal{K}) - \Lambda_{\mathcal{C}}^{y_i}(\mathcal{K}) \geq \delta,\end{aligned}$$

where $\Lambda_{\mathcal{C}}^{y_i}(G)$ represents the prediction accuracy of collective prediction method $\mathcal{C}$ for NSLA y_i. ϵ is the super-threshold of social structure changes. δ measures how much added prediction accuracy is earned by adversaries through predicting with the published G'. Clearly, $\epsilon, \delta \geq 0$. To preserve data utility, both ϵ and δ are given by the data publisher.

Next, we define utility loss due to the data-sanitization strategy carried out on published data. Definition 5.1 shows that utility loss comes from two aspects: network structure disparity and prediction accuracy deviation for NSLA. Therefore, utility loss is defined based on the above two aspects: structure utility loss and prediction utility loss.

Definition 5.2. ϵ-Structure utility loss: Structure utility loss estimates how much an arbitrary user u_i loses regarding network structure after sanitizing its social links. Structure utility loss of u_i is determined by the structure utility values of u_i's neighbors. For a given structure utility value metric, the ϵ-structure utility loss for u_i after sanitizing u_i' vulnerable link set $A_i \subseteq N_i$ is given by $SUL_i = \zeta(\mathcal{S}_{A_i}) \leq \epsilon$, where $\mathcal{S}_{A_i} = \{S_j | u_j \in A_i \subseteq N_i, S_j \in \mathbb{R}^*\}$, and S_j represents the structure utility value of user u_j.

The structure utility value of a user reflects social structure properties, which can be measured by different metrics. In this paper, we use number of shared friends as structure utility metric. Unfriending a friend that shares a large of friends of one user has a bad effect on the clustering coefficient of the user [ ]. Furthermore,we assume $\zeta(.)$ is an additive function,then $SUL_i = \sum_{u_j \in A \subseteq N_i} S_j \leq \epsilon$.

Since both attribute set and social links of a user are sanitized and we assume powerful adversaries predict SLA based on them simultaneously as shown in Section IV-A, prediction utility loss is derived from both of the disparity sources. Since social structure disparity is measured by ϵ-structure utility loss, prediction utility loss only needs to measure the prediction accuracy deviation derived from attribute sanitization.

To evaluate prediction utility loss due to sanitized attribute set, we introduce an attribute set disparity measurer d_u, such that $d_u(X, X')$ measures how much prediction utility loss there is if a SNA server performs analysis depending on X' rather than X. Thus, given $\psi(X)$, $f(X'|X)$, and $d_u(X, X')$, prediction utility loss can be calculated as the expectation of $d_u(X, X')$ over all X and X' for a user.

Definition 5.3. δ-Prediction utility loss: Prediction utility loss estimates the amount of prediction accuracy deviation for the NSLA of an arbitrary user u_i. For a given attribute set disparity measurer d_u, the δ-prediction utility loss for u_i after carrying out a data sanitization method on its attribute set X and social links, is given by $PUL_i = \sum_{X,X'} \psi(X) f(X'|X) d_u (X, X') \leq \delta$.

Attribute set disparity measurer d_u is determined by data semantics. In different applications, d_u can be defined as Euclidean, Hamming, or Mahalanobis distance, etc.

B. Latent-Data Privacy Metric

We assume powerful adversaries have the knowledge of user's profile $\psi(X)$ and our data-sanitization strategy. After obtaining the sanitized attribute set, adversaries calculate the posterior probability of X, conditional on X' with prior knowledge $\psi(X)$ and $f(X'|X)$:

$$Pr(X|X') = \frac{Pr(X, X')}{Pr(X')} = \frac{f(X'|X)\psi(X)}{\sum_X f(X'|X)\psi(X)}$$

Then, for each X with posterior probability $Pr(X|X')$, adversaries can predict the user's SLA based on X and sanitized social links. We represent the SLA predicted from $Pr(X|X')$ as Z_X. Obviously, Z_X is related to the sanitized link set A such that we denote Z_X as the function of A, *i.e.*, $Z_X(A)$. Adversaries' goal is then to choose $\hat{Z}$ to minimize the user's conditional expected latent-data privacy, conditional on $Pr(X|X')$. For an arbitrary $\hat{Z}$, the user's conditional expected latent-data privacy is $\sum_X Pr(X|X') d_p(Z_X(A), \hat{Z})$, where $d_p(Z_X(A), \hat{Z})$ is the privacy disparity between $Z_X(A)$ and $\hat{Z}$.

For the minimized $\hat{Z}$, it is

$$\min_{\hat{Z}} \sum_X Pr(X|X') d_p(Z_X(A), \hat{Z}) \tag{4}$$

The latent-data privacy conditional on a given X' is given by Equation (4). Meanwhile, the probability of X' output by the sanitization method is $P(X') = \sum_X f(X'|X)\psi(X)$. Thus, unconditional expected privacy of the user's is

$$\sum_{X'} \psi(X') \min_{\hat{Z}} \sum_{X} Pr(X|X') d_p(Z_X(A), \hat{Z})$$
$$= \sum_{X'} \min_{\hat{Z}} \sum_{X} \psi(X) f(X'|X) d_p(Z_X(A), \hat{Z}) \quad (5)$$

We define

$$P_{X'} = \min_{\hat{Z}} \sum_{X} \psi(X) f(X'|X) d_p(Z_X(A), \hat{Z}). \quad (6)$$

Incorporating $P_{X'}$ into Equation (5), the users unconditional expected privacy is rewritten as

$$\sum_{X'} P_{X'}, \quad (7)$$

which is the user attempts to maximize by finding the optimal $f(X'|X)$.

Unfortunately, the minimum operator in Equation (6) makes the computation problem nonlinear. Therefore, we can transform (6) into a series of linear constraints by

$$P_{X'} \leq \sum_{X} \psi(X) f(X'|X) d_p(Z_X(A), \hat{Z}) \quad \forall \hat{Z} \quad (8)$$

Therefore, maximizing Formula (7) under constraint (6) is equal to optimizing Formula (7) under constraint (8).

VI. Privacy-Utility Tradeoff

In this section, we first formalize optimal problem that can produce optimized data sanitization strategy. Then, we discuss how to solve the proposed optimal problem. Here, we introduce function *LaPri*(.) to measure latent-data privacy with current sanitized attribute set and social links.

A. Optimal Problem Formulation

The problem of (ϵ, δ)-utility with maximize latent-data privacy can be formulated as follows.

Definition 6.1. (ϵ, δ)*-UtiOptPri* $(\psi(.), d_u(.), d_p(.), \mathcal{S}, \epsilon, \delta,)$*:* Given user's profile $\psi(.)$, attribute set disparity measurer $d_u(.)$, privacy disparity measure $d_p(.)$, structure utility value metric $\mathcal{S}$, structure utility loss threshold ϵ, and prediction utility loss threshold δ, the question is to find data-sanitization strategy $f(.)$ and link-sanitization strategy $\mathcal{L}(.)$, and latent-data privacy function LaPri(.) such that

1) $\mathcal{L}(.)$ satisfies ϵ-structure utility loss and $f(.)$ satisfies δ-prediction utility loss;
2) For any $\mathcal{L}'(.)$ that satisfies ϵ-structure utility loss and $f'(.)$ that satisfies δ-prediction utility loss, $LaPri(\mathcal{L}'(.), f'(.), \psi(.), d_p(.)) \geq LaPri(\mathcal{L}(.), f(.), \psi(.), d_p(.))$.

The linear optimization program for an arbitrary user u_i to find the optimal data sanitization strategy is as following: choose $f(X'|X), \hat{Z}, \forall X, X'$, in order to

$$\textbf{Maximize:} \sum_{X'} P_{X'}$$

$$\textbf{Subject to}:$$
$$P_{X'} \leq \sum_{X} \psi(X) f(X'|X) d_p(Z_X(A), \hat{Z}) \quad \forall \hat{Z}$$
$$\sum_{u_j \in A_i \subseteq N_i} S_j \leq \epsilon$$
$$\sum_{X} \psi(X) \sum_{X'} f(X'|X) d_u(X, X') \leq \delta$$
$$f(X'|X) \geq 0 \quad \forall X, X'$$
$$\sum_{X'} f(X'|X) = 1, \ \forall X$$

B. Solve the Optimal Problem

We now solve the optimal problem to find attribute sanitization strategy $f(.)$ and link sanitization strategy $\mathcal{L}(.)$.

B1) Find Link-Sanitization Strategy: First, we prove the link sanitization method introduced in Section IV-B2 has monotonicity property. The monotonicity property indicates that if we increase the number of removed links of a user, we can only improve this user's latent-data privacy.

Theorem 6.1. Monotonicity: Function $LaPri : A_i \rightarrow \mathbb{R}^*$ is monotonically nondecreasing, namely, $LaPri(A_i \cup e_{ij}) \leq LaPri(A_i)$, where $e_{ij} \in A_i$, $A_i \in N_i$, and A_i is the vulnerable link set of u_i.

Proof: As discussed in Definition 4.1, the prediction accuracy for user u_i' SLA decreases upon removing the vulnerable link between u_i and u_j; namely, for any vulnerable link e_{ij}, $\Lambda(A_i) \leq \Lambda(A_i \cup e_{ij})$. This accuracy relation indicates that for user u_i, the latent-data privacy with vulnerable link set A_i is definitely larger than the latent-data privacy with vulnerable link set $A_i \cup e_{ij}$. Hence, $LaPri(A_i \cup e_{ij}) \leq LaPri(A_i)$. ■

Theorem 6.2. Submodularity: Function $LaPri : A_i \rightarrow \mathbb{R}^*$ is submodular, namely, $LaPri(B_i \cup e_{ij}) - LaPri(B_i) \leq LaPri(A_i \cup e_{ij}) - LaPri(A_i)$, where $A_i \subseteq B_i \subseteq N_i$, $e_{ij} \in N_i$, and A_i and B_i are vulnerable link sets of u_i.

Proof: For the prediction accuracy for SLA, the maximum decrease in prediction accuracy of user u_i, by removing a vulnerable link e_{ij} from vulnerable link set A_i is at least more than the maximum decrease by removing e_{ij} from another set B_i, namely, $\Lambda(A_i \cup e_{ij}) - \Lambda(A_i) \leq \Lambda(B_i \cup e_{ij}) - \Lambda(B_i)$, where $A_i \subseteq B_i \subseteq N_i$, and $e \in N_i$. The accuracy relation indicates that for user u_i, the maximum gain in latent-data privacy after removing vulnerable link e_{ij} from vulnerable link set A_i is at least more than the maximum gain by removing e_{ij} from B_i. Hence, $LaPri(B_i \cup e_{ij}) - LaPri(B_i) \leq LaPri(A_i \cup e_{ij}) - LaPri(A_i)$. ■

With Theorems 6.1 and 6.2, the problem of finding a link-sanitization strategy is equivalent to the minimization of submodular, nondecreasing, nonnegative function with constraints that is knapsack-like. The greedy algorithm proposed in [30]

TABLE II
GENERAL INFORMATION ABOUT CALTECH

Network property	Value
Number of users	769
Number of social links	16656
Number of attributes of each user	7
Number of possible attribute values for SLA	4
Number of possible attribute values for NSLA	2

could be exploited to solve this problem with nondecreasing, submodular, nonnegative objective function constrained by structure utility loss.

B2) Find Attribute-Sanitization Strategy: To find an attribute-sanitization strategy, the optimization problem can be solved by iterating over all possible $f(X'|X)$, all X and all sanitized X' to make sure the prediction accuracy loss of latent-data privacy is less than δ. Furthermore, find the optimal set of $f(X'|X)$ that produce minimum value of objective function $\sum_{X'} P_{X'}$. However, this approach is impractical since there is an infinite number of $f(X'|X)$. For example, X has three possible sanitized attribute vectors X_1, X_2 and X_3, and the probabilities that satisfy $\sum_{i=1:3} f(X_i|X) = 1$, $f(X_i|X) \geq 0, \forall X, X_i$ are infinite. To solve this problem, we discrete the probability space, *i.e.*, $[0, \ldots, 1] \rightarrow [0, 1/d, 2/d, \ldots, 1]$ to get a suboptimal solution. Furthermore, to shrink search space of X, the set of X' can be derived through substituting each attribute in the shared attributes between SLA and NSLA with a generic attribute, which ensures that adversaries cannot get specific information to increase prediction accuracy on sensitive attributes, while guarantees no significant accuracy reduction on data utility. Moreover, since there are different levels of generalization, we organize the generic attributes as a hierarchy.

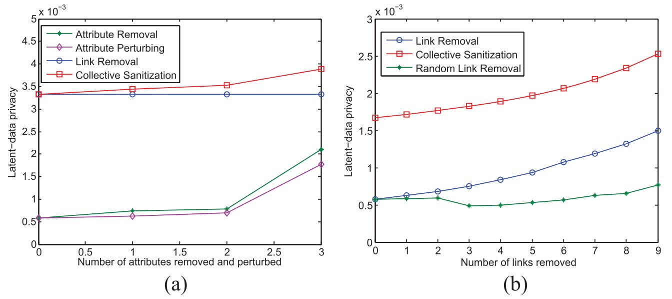
VII. EVALUATION

A. Dataset

In our evaluation, we study a large Facebook dataset that contains all the Facebook "friendship" links among the users at California Institute of Technology at a certain time in September 2005. It also includes some demographic information like student/faculty status flag, gender and some other attributes, which are published by users on their Facebook pages. Each attribute is assigned a numeric value and user identity is ignored. For convenience, the dataset is named as *Caltech*. Some general information about Caltech are listed in Table II.

B. Experimental Settings

As shown in Table II, there are 7 attributes for each user. We choose attribute *student/faculty status flag* (represented by *flag*) and *gender* as SLA and NSLA, respectively. Table II shows that SLA and NSLA have 4 and 2 possible attribute values, respectively. The remaining 5 attributes are assumed to be publicly available attributes, among which 3 attributes are for SLA, 3 attributes are for NSLA, and 1 attribute is common.



(a) (b)

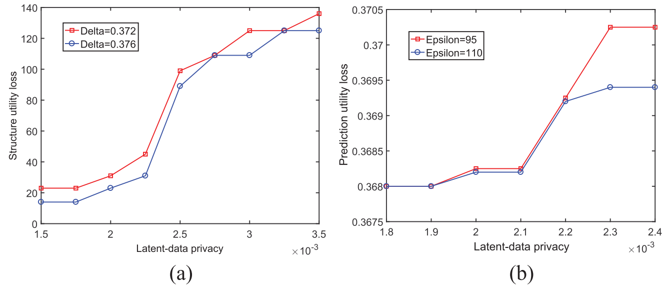
Fig. 1. Latent-data privacy under different data-sanitization strategies with increasing number of (a) attributes; (b) sanitized links, $\epsilon = 180$, and $\delta = 0.4$.

We compare our data-sanitization strategy with different strategies to satisfy the (ϵ, δ)-UtiOptPri problem defined in Definition 6.1: 1) Attribute Removal: remove the most indicative attributes for SLA; 2) Attribute Perturbing: perturb the most indicative attributes for SLA; 3) Link Removal: remove vulnerable links; 4) Random Link Removal: randomly remove links. We denote our data sanitization strategy as *Collective Sanitization* since it collectively harnesses different data sanitization methods.

C. Privacy-Utility Tradeoff With Different Data-Sanitization Strategies

We evaluate the effectiveness of our Collective Sanitization to realize the privacy-utility tradeoff. To make a fair comparison, we first evaluate latent-data privacy when the above five strategies satisfy the same data utility thresholds. We choose an arbitrary pair of ϵ and δ such as $\epsilon = 180$, $\delta = 0.4$, and then calculate the latent-data privacy under different strategies with increasing number of attributes and links being sanitized. As stated in Section IV-B1, Collective Sanitization sanitizes user attributes by employing removing and perturbing collectively. The horizontal axis of Fig. 1(a) for Collective Sanitization represents the number of the removed attributes (indicative for SLA) and the number of attributes (common indicative attributes for SLA and NSLA) being perturbed. Similarly, the horizontal axis of Fig. 1(b) for Collective Sanitization represents the number of the removed vulnerable links (as presented in Section IV-B2).

As shown in Fig. 1(a), four strategies are generally effective in protecting latent-data privacy while realizing customized (ϵ, δ)-utility. With increasing number of attributes being sanitized, latent-data privacy monotonically increases as well. However, compared with the remaining three strategies, *Collective Sanitization* can realize a larger level of latent-data privacy with the same number of attributes being sanitized and same utility thresholds. Meanwhile, as expected, *Attribute Removal* is better than *Attribute Perturbing* in protecting latent-data privacy. With more and more attributes removed and perturbed, this advantage of *Attribute Removal* becomes more and more obvious. Furthermore, in protecting latent-data privacy, *Link Removal* is better than both *Attribute Removal* and *Attribute Perturbing*. To explain this observation, we find that the latent-data privacy under



(a) (b)

Fig. 2. Utility loss under different levels of latent-data privacy. (a) Structure utility loss with different prediction utility loss thresholds and $\epsilon = 180$. (b) Prediction utility loss with different structure utility loss thresholds and $\delta = 0.4$.

Link Removal and *Collective Sanitization* are close, indicating that removing vulnerable links contributes more effectiveness than attribute sanitization in protecting latent-data privacy.

The same observation can be found in Fig. 1(b), where latent-data privacy monotonically increases with more and more links removed. However, compared with the remaining two strategies, *Collective Sanitization* can also achieve a larger level of latent-data privacy with the same number of links removed and same utility threshold. Meanwhile, *Link Removal* is better than *Random Link Removal* in protecting latent-data privacy. With more and more links removed, this advantage of *Link Removal* becomes more and more obvious.

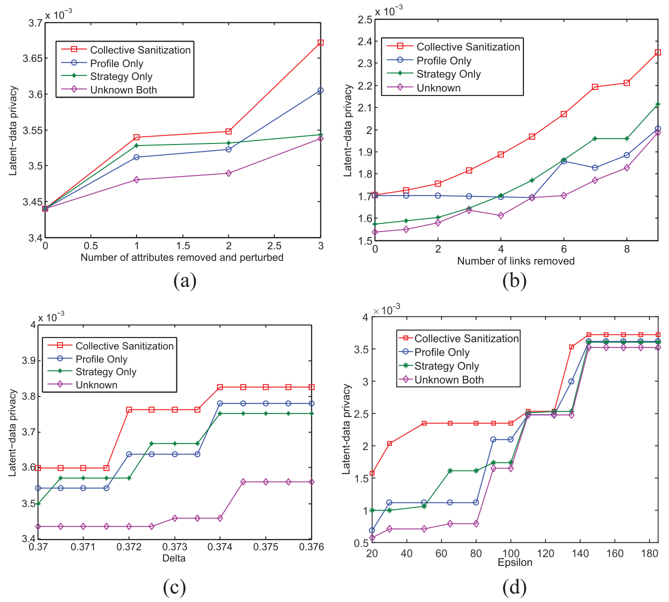
We further discuss the effectiveness of Collective Sanitization in guaranteeing utility under different levels of latent-data privacy. The results are shown in Fig. 2, which shows that utility loss increases with the increasing of latent-data privacy level. Furthermore, utility loss converges to a stable level with the increasing of latent-data privacy level. The reason lies that the marginal gain of latent-data privacy is obtained with the maximum number of sanitized attributes and links, and minimized utility.

D. Privacy-Utility Tradeoff With Different Prior Knowledge

We evaluate the privacy-utility tradeoff with different cases of prior knowledge for adversaries. We compare our Collective Sanitization assuming most powerful adversaries with the knowledge of user profile $\psi(X)$ and data-sanitization strategy, where different types of prior knowledge are assumed: 1) Profile Only: only profile is known to adversaries; 2) Strategy Only: only data-sanitization strategy is known to adversaries; 3) Unknown Both: neither profile nor strategy is known to adversaries.

To make a fair comparison, we first compare the latent-data privacy when the above four cases has same utility thresholds. With the same utility thresholds $\epsilon = 500$ and $\delta = 0.4$, we calculate the latent-data privacy under different cases with increasing number of sanitized attributes and links. The results are shown in Fig. 3(a) and (b), where the horizontal axis for Collective Sanitization represents the number of removed/perturbed attributes and the number of removed vulnerable links, respectively.

Fig. 3 shows that compared with different cases, Collective Sanitization assuming powerful adversaries is the most effective one in protecting latent-data privacy while guaranteeing customized (ϵ, δ)-utility. As shown in Fig. 3(a) and (b), the latent-data privacy under Profile Only and Strategy Only lies somewhere in between Collective Sanitization and Unknown Both, and profile information is more valuable than strategy information in some cases. The similar observation can be obtained in Fig. 3(c) and (d), where it is also shown that latent-data privacy converges to a stable level with the increasing of utility thresholds. The reason lies that the marginal gain of latent-data privacy is obtained with the most sacrifice in utility.



(a) (b) (c) (d)

Fig. 3. Latent privacy-utility tradeoff with different cases of prior knowledge for adversaries, with increasing number of (a) attributes; (b) sanitized links; and the increasing of (c) prediction utility threshold; (d) structure utility threshold.

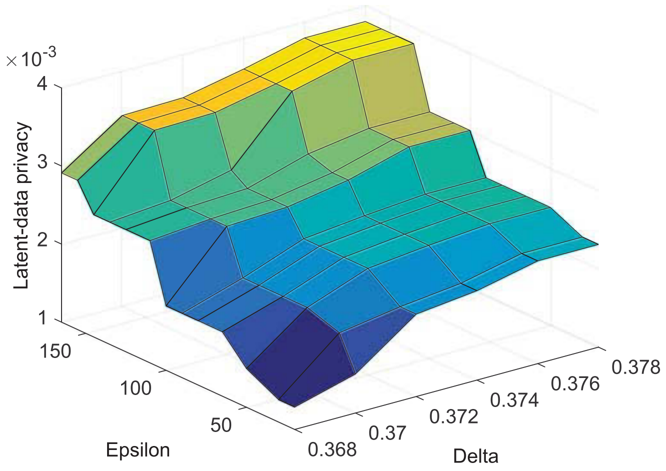


Fig. 4. Latent-data privacy with different utility thresholds.

Finally, the latent-data privacy with different utility thresholds is shown in Fig. 4. Fig. 4 shows that with the increasing of ϵ and δ, latent-data privacy increases as well. The reason lies that it is possible to determine a better data-sanitization strategy with fewer utility requirements. Furthermore, latent-data privacy converges to a stable value with continuously increase of ϵ and δ, which indicates the optimal data-sanitization strategy is found.

VIII. Conclusion

In this paper, we study how to optimize the tradeoff between latent-data privacy and customized data utility when combating against powerful adversaries with optimal inference attacks. To address this issue, we first propose two sanitization methods for links and attributes, based on which we formalize prediction utility loss matric, structure utility loss matric and latent-data privacy. Then we formulate an optimization problem that can maximize latent-data privacy while guaranteeing customized data utility. Finally, we evaluate our data-sanitization strategy towards real big social network data and the results show that the proposed data-sanitization strategy can effectively achieve a meaningful privacy-utility tradeoff. Our future work is to explore formal privacy models, such as differential privacy or k-anonimity to balance latent-data privacy and customized data utility.

References

[1] C. Y. Johnson, "Project Gaydar," *Boston Globe*, Sep. 2009.

[2] N. Z. Gong *et al.*, "Joint link prediction and attribute inference using a social-attribute network," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 2, Apr. 2014, Art. no. 27.

[3] Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla, "Inferring user demographics and social strategies in mobile social networks," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 15–24.

[4] S. Volkova, Y. Bachrach, M. Armstrong, and V. Sharma, "Inferring latent user properties from texts published in social media," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 4296–4297.

[5] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in online social networks," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, 2010, pp. 251–260.

[6] R. Heatherly, M. Kantarcioglu, and B. M. Thuraisingham, "Preventing private information inference attacks on social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 8, pp. 1849–1862, Aug. 2013.

[7] Z. He, Z. Cai, and Y. Li, "Customized privacy preserving for classification based applications," in *Proc. 1st ACM Workshop Privacy-Aware Mobile Comput.*, 2016, pp. 37–42.

[8] Z. Jorgensen, T. Yu, and G. Cormode, "Publishing attributed social graphs with formal privacy guarantees," in *Proc. Int. Conf. Manage. Data*, 2016, pp. 107–122.

[9] J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava, and X. Xiao, "Privbayes: Private data release via Bayesian networks," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2014, pp. 1423–1434.

[10] C. Liu and P. Mittal, "Linkmirage: Enabling privacy-preserving analytics on social relationships," in *Proc. Netw. Distrib. Syst. Security Symp.*, 2016, pp. 492–503.

[11] W.-Y. Day, N. Li, and M. Lyu, "Publishing graph degree distribution with node differential privacy," in *Proc. Int. Conf. Manage. Data*, 2016, pp. 123–138.

[12] P. Gundecha, G. Barbier, J. Tang, and H. Liu, "User vulnerability and its reduction on a social networking site," *ACM Trans. Knowl. Discovery Data*, vol. 9, no. 2, Sep. 2014, Art. no. 12.

[13] M. Han, M. Yan, Z. Cai, and Y. Li, "An exploration of broader influence maximization in timeliness networks with opportunistic selection," *J. Netw. Comput. Appl.*, vol. 63, pp. 39–49, 2016.

[14] Z. He, Z. Cai, Q. Han, W. Tong, L. Sun, and Y. Li, "An energy efficient privacy-preserving content sharing scheme in mobile social networks," *Pers. Ubiquitous Comput.*, vol. 20, no. 5, pp. 833–846, 2016.

[15] Z. He, Z. Cai, Y. Sun, Y. Li, and X. Cheng, "Customized privacy preserving for inherent data and latent data," *Pers. Ubiquitous Comput.*, vol. 21, no. 1, pp. 43–54, Feb. 2017.

[16] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Trans. Dependable Secure Comput.*, to be published.

[17] Z. He, Y. Li, J. Li, J. Yu, H. Gao, and J. Wang, "Addressing the threats of inference attacks on traits and genotypes from individual genomic data," in *Proc. Int. Symp. Bioinformat. Res. Appl.*, 2017, pp. 223–233.

[18] X. Zheng, Z. Cai, J. Li, and H. Gao, "A study on application-aware scheduling in wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 7, pp. 1787–1801, Jul. 2017.

[19] X. Zheng and Z. Cai, "Real-time big data delivery in wireless networks: A case study on video delivery," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 2048–2057, Aug. 2017.

[20] X. Zheng, Z. Cai, J. L, and H. G, "Location-privacy-aware review publication mechanism for local business service systems," in *Proc. 36th Annu. IEEE Int. Conf. Comput. Commun.*, 2017.

[21] T. Qiu, R. Qiao, and D. Wu, "EABS: An event-aware backpressure scheduling scheme for emergency internet of things," *IEEE Trans. Mobile Comput.*, to be published.

[22] T. Qiu, A. Zhao, R. Ma, V. Chang, F. Liu, and Z. Fu, "A task-efficient sink node based on embedded multi-core SOC for internet of things," *Future Gener. Comput. Syst.*, 2016.

[23] Y. E. Sun *et al.*, "Privacy-preserving strategyproof auction mechanisms for resource allocation," *Tsinghua Sci. Technol.*, vol. 22, no. 2, pp. 119–134, Apr. 2017.

[24] J. Lu and X. Wang, "Interference-aware probabilistic routing for wireless sensor networks," *Tsinghua Sci. Technol.*, vol. 17, no. 5, pp. 575–585, 2012.

[25] Z. He, Y. Li, and J. Wang, *Differential Privacy Preserving Genomic Data Releasing via Factor Graph*. New York, NY, USA: Springer, 2017, pp. 350–355.

[26] Z. He, Z. Cai, J. Yu, X. Wang, Y. Sun, and Y. Li, "Cost-efficient strategies for restraining rumor spreading in mobile social networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2789–2800, Mar. 2017.

[27] C. Dwork, "Differential privacy," in *Automata, Languages and Programming (ser. Lecture Notes in Computer Science)*, vol. 4052, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, Eds. Berlin, Germany: Springer, 2006, pp. 1–12.

[28] L. Sweeney, "K-anonymity: A model for protecting privacy," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, Oct. 2002.

[29] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, Mar. 2007, Art. no. 3.

[30] M. Sviridenko, "A note on maximizing a submodular set function subject to a knapsack constraint," *Oper. Res. Lett.*, vol. 32, no. 1, pp. 41–43, 2004.

Zaobo He (S'14) received the M.S. degree in computer science from Shannxi Normal University, Xi'an, China, in 2013. He is currently working toward the Ph.D. degree with the Department of Computer Science, Georgia State University, Atlanta, GA, USA. His research interests include privacy preservation, social networking, and big data.

Zhipeng Cai (SM'14) received the B.S. degree from Beijing Institute of Technology, Beijing, China, and the M.S. and Ph.D. degrees in computing science from the University of Alberta, Edmonton, AB, Canada. He is currently an Assistant Professor with the Department of Computer Science, Georgia State University, Atlanta, GA, USA. His research interests include networking, privacy, and big data. He has received an NSF CAREER Award. He is an Editor/Guest Editor for *Algorithmica*, *Theoretical Computer Science*, *Journal of Combinatorial Optimization*, and the IEEE/ACM Transactions on Computational Biology and Bioinformatics.

Jiguo Yu received the Ph.D. degree in mathematics from Shandong University, Jinan, China, in 2004. Since 2007, he has been a Professor with the School of Computer Science, Qufu Normal University, Rizhao, China, where he is also a Professor with the School of Information Science and Engineering. His main research interests include wireless networks, distributed algorithms, peer-to-peer computing, and graph theory. He is a senior member of the China Computer Federation.