The Privacy-Utility Tradeoff in Rank-Preserving Dataset Obfuscation

Mahshad Shariatnasab, Farhad Shirani, S. Sitharma Iyengar
Florida International University
Email: mshar075@fiu.edu, fshirani@fiu.edu, iyengar@fiu.edu

Abstract

Dataset obfuscation refers to techniques in which random noise is added to the entries of a given dataset, prior to its public release, to protect against leakage of private information. In this work, dataset obfuscation under two objectives is considered: i) rank-preservation: to preserve the row ordering in the obfuscated dataset induced by a given rank function, and ii) anonymity: to protect user anonymity under fingerprinting attacks. The first objective, rank-preservation, is of interest in applications such as the design of search engines and recommendation systems, feature matching, and social network analysis. Fingerprinting attacks, considered in evaluating the anonymity objective, are privacy attacks where an attacker constructs a fingerprint of a victim based on its observed activities, such as online web activities, and compares this fingerprint with information extracted from a publicly released obfuscated dataset to identify the victim. By evaluating the performance limits of a class of obfuscation mechanisms over asymptotically large datasets, a fundamental trade-off is quantified between rank-preservation and user anonymity. Single-letter obfuscation mechanisms are considered, where each entry in the dataset is perturbed by independent noise, and their fundamental performance limits are characterized by leveraging large deviation techniques. The optimal obfuscating test-channel, optimizing the privacy-utility tradeoff, is characterized in the form of a convex optimization problem which can be solved efficiently. Numerical simulations of various scenarios are provided to verify the theoretical derivations.

This work was supported in part by NSF grants CCF-2241057.

I. Introduction

Dataset privacy is a major concern due to the potential risks associated with the misuse of personal and sensitive information included in various datasets. If the data to be released has no immediate utility, then cryptographic methods suffice to preserve privacy [1], [2]. However, when data is released publicly for a specific immediate utility — such as the release of *anonymized* social network data to advertising companies — the necessarily unencrypted disclosure incurs a privacy risk and may lead to unwanted inferences [3]–[8]. Obfuscation provides a mitigating solution, by introducing noise in the dataset entries prior to their release. This leads to a privacy-utility tradeoff, where increased perturbation of the dataset entries via random noise leads to increased privacy at the expense of lost utility. In this work, we study this fundamental privacy-utility tradeoff and characterize optimal obfuscation strategies, where privacy is evaluated under fingerprinting attacks [5]–[8], and utility is measured via metrics associated with rank-preservation [9]–[12].

Obfuscation mechanisms protect privacy via noisy perturbations of the dataset entries. A widely studied class of obfuscation mechanisms is to perturb each dataset entry independently by passing them through identical test-channels [5], [13]–[15]. We call these mechanisms *single-letter obfuscation mechanisms* since their operations can be characterized using single-letter conditional probability measures. Single-letter obfuscation mechanisms, as opposed to multi-letter mechanisms, are amiable to analysis, and they have good performance under specific utility metrics such as the variational distance and Euclidean distance metrics [16]–[19]. Furthermore, perturbation via independent noise reduces information leakage among entries of the obfuscated dataset. Consequently, in this work, we focus our study to single-letter obfuscation mechanisms and their fundamental performance limits.

Rank-preservation is a utility metric of interest in dataset obfuscations [9]–[12], [20], [21]. In general, for a given dataset X with $n \in \mathbb{N}$ rows, a rank function $R : [n] \to [n]$ is a mapping which assigns an ordering to the rows of the dataset. For instance, let us consider a social network with $n \in \mathbb{N}$ users, and let $\mathsf{X} = [X_{i,j}]_{i,j \in [n]}$ be the adjacency matrix capturing the user's connections in the social network, where $X_{i,j} = 1$ if the ith and jth users are connected, and

$X_{i,j} = 0$ otherwise. The user-degree-based rank function induces an ordering of the users based on number of connections, i.e. $R(i) < R(i')$ if $\sum_{j\in[n]} X_{i,j} < \sum_{j\in[n]} X_{i',j}$. Rank functions are used in the design of search engines, social network analysis, feature matching, and recommendation systems [9], [22], [23]. Rank recovery algorithms reconstruct the rank function associated with a given dataset based on noisy observations, e.g., by observing an obfuscated dataset. That is, given an obfuscated dataset Y, a rank-recovery algorithm produces a reconstruction $\widehat{R}(\cdot)$ of the rank function $R(\cdot)$ associated with the original dataset X. The performance of the rank recovery algorithm is measured with respect to an underlying distortion metric, measuring the distance between the original and recovered rank-functions. A widely used distortion metric, considered in this work, is the Kendall's rank correlation coefficient (KRCC) [9], [10], [16], [23]. The KRCC distance $d(R,\widehat{R})$ counts the number of pairwise disagreements between the two rank functions, i.e. $d(R,\widehat{R}) \triangleq \sum_{i\in[n]} \mathbb{1}(R(i) \neq \widehat{R}(i))$.

We study the privacy-utility tradeoff in database obfuscation, where the utility objective is rank-preservation discussed in the prequel, and privacy is evaluated under fingerprinting attacks. Fingerprinting attacks are a major threat to users' privacy in social networks, mobility networks, and wireless networks, among others [24]–[26]. In these attacks, given an obfuscated dataset, the attacker's objective is to identify the row in the dataset corresponding to a victim by acquiring a partial fingerprint of the victim's real-world activities, comparing it with each of the rows in the obfuscated dataset, and detecting the row with correlated entries (Figure 1). To provide an example, let us consider online fingerprinting attacks which rely on social network group memberships [6]–[8]. In such scenarios, an attacker controls a malicious website, the victim is a visitor to the website, and the attacker uses browser history sniffing techniques to extract a partial list of social network groups visited by the victim [4], [8]. The extracted information can be represented by a binary vector $F^q = (F_1, F_2, \cdots, F_q), q \in \mathbb{N}$, where $F_i = 1$ if the victim has visited the ith social network group's website, and $F_i = 0$, otherwise. The vector F^q is called the fingerprint of the victim. To identify the victim's social network account, the attacker scans the social network and acquires a (obfuscated) dataset Y capturing the public group memberships in the social network. It then compares the fingerprint F^q and the dataset Y to find the closest match

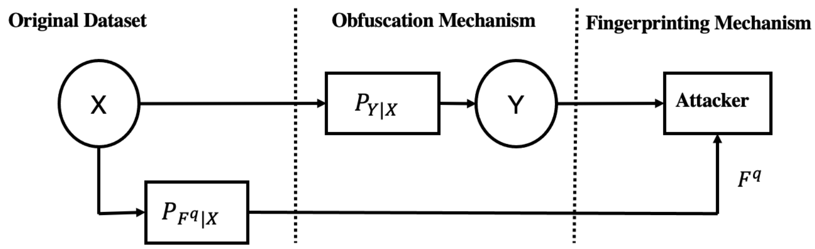

Fig. 1. X represents the original dataset. Y represents the obfuscated dataset. The attacker acquires the fingerprint vector F^q by querying the victim's activities and attempts to identify the victim by comparing F^q and Y.

and identify the victim. In practice, the attacker acquires each fingerprint element by *querying* the user's activities, and there is a cost associated with each query. For instance, in social network fingerprinting attacks described above, the state-of-the-art browser history sniffing techniques can make between tens to several thousand queries per second depending on the victim's device and web browser [27], [28]. So, the cost associated with each fingerprint element is the time spent to query the value of that element using browser history sniffing. As a result, the length of the partial fingerprint is determined by the attacker's resources. In this work, the privacy objective under consideration is to minimize the information leakage about the victim's identity given a partial fingerprint F^q with a fixed length $q \in \mathbb{N}$.

The following is a summary of our contributions:

- To formulate the dataset obfuscation problem under the rank-preservation and anonymity constraints.
- To evaluate the fundamental performance limits of single-letter obfuscation mechanisms and quantify a tradeoff between the two objectives. This allows the system designer to choose the appropriate amount of obfuscation through the choice of a single-letter test-channel by optimizing the aforementioned trade-off.
- To characterize the optimal obfuscating test-channel, optimizing the privacy-utility tradeoff, in the form of a convex optimization problem.
- To provide numerical simulations under various statistical scenarios.

Notation: The random variable $\mathbb{1}_{\mathcal{E}}$ is the indicator of the event $\mathcal{E}$. The set $\{n, n+1, \cdots, m\}, n, m \in$

$\mathbb{N}$ is represented by $[n,m]$, and for the interval $[1,m]$, we use the shorthand notation $[m]$. For a given $n \in \mathbb{N}$, the n-length vector $(x_1, x_2, \ldots, x_n)$ is written as x^n, $\underline{x}$, and $(x_i)_{i\in[n]}$, interchangably. The notation $[x_{i,j}]_{i\in[n],j_[m]}$ denotes an $n \times m$ matrix, where $x_{i,j}$ is the element on the ith row and jth column. We use sans-serif letter such as X and x to represent matrices.

II. Problem formulation

In this section, we describe the mathematical formulation of the dataset obfuscation problem shown in Figure 2.

Random Dataset: A dataset is a matrix $\mathsf{X} = [x_{i,j}]_{i\in[n],j\in[m]}$, where $x_{i,j} \in \mathcal{X}$, the set $\mathcal{X}$ is finite, and $n, m \in \mathbb{N}$. Each row $x_i^m = (x_{i,1}, x_{i,2}, x_{i,3}, \cdots x_{i,m}), i \in [n]$ is called an entry of the dataset, $m \in \mathbb{N}$ is the length of the entries, $n \in \mathbb{N}$ is the size of the dataset. The dataset is said to have n members. We consider stochastically generated datasets with independent and identically distributed (IID) elements, where given a distribution P_X defined on alphabet $\mathcal{X}$, we have:

$$P(\mathsf{X} = [x_{i,j}]_{i\in[n],j\in[m]}) = \prod_{i\in[n],j\in[m]} P_X(x_{i,j})$$

A random dataset is parameterized by $(n, m, \mathcal{X}, P_X)$.

Original and Obfuscated Datasets: An agent, Alice, has access to an original dataset X parameterized by $(n, m, \mathcal{X}, P_X)$. Alice wishes to disclose an obfuscated dataset $\mathsf{Y} = f(\mathsf{X})$ to Bob, where $f(\cdot)$ is a possibly stochastic function captured by $P_{\mathsf{Y}|\mathsf{X}}$. Bob's objective is to recover the row-ordering of the original dataset, with respect to a given rank function $R(\cdot)$, by leveraging the obfuscated dataset. The rank function and privacy constraints under consideration are described in more detail in the sequel.

Privacy Objective: An attacker, Eve, gains access to the disclosed dataset Y. Eve's objective is to identify the dataset entry corresponding to a specific victim. To elaborate, we let U represent the row index corresponding to the victim of interest. The random variable U is assumed to be uniformly distributed on $[n]$. Eve acquires a partial fingerprint F^q of the row elements $(X_{U,i_1}, X_{U,i_2}, \cdots, X_{U,i_q})$ corresponding to the victim, where

$$P(F^q = f^q | (X_{U,i_j})_{j\in[q]} = x^q) = \prod_{i=1}^{q} P_{F|X}(f_i | x_{i_j}), f^q, x^q \in \mathcal{F}^q \times \mathcal{X}^q,$$

and $P_{F|X}$ is a collection of $|\mathcal{X}|$ probability measures defined on a finite set $\mathcal{F}$. The fingerprint vector and obfuscated dataset are conditionally independent of each other given the original dataset, i.e., the Markov chain $F_i \leftrightarrow X_{U,i_j} \leftrightarrow Y_{U,i_j}, i_j \in [m], j \in [q]$ holds. One of the objectives in the dataset obfuscation problem is to minimize the information leakage between the victim's row index and Eve's observations (Y, F^q). That is to minimize $I(U; \mathsf{Y}, F^q)$. We assume that the fingerprinting process is unsupervised in the sense that Eve does not have a choice on which indices $i_j, j \in [q]$ are queried to extract the fingerprint.

Rank-Preservation Objective: In general, given a dataset X a rank function is a mapping $R : [n] \to [n]$ which induces an ordering on the rows in the dataset, i.e., $R(i)$ indicates the rank of the ith row of X induced by the rank function $R(\cdot)$. In this work, we consider the degree-based rank function defined in the following. The degree-based rank function and its variants are used in applications such as social network analysis, search engine design, and recommendation systems [29]–[31].

Definition 1 (**Degree-Based Rank Function**). Given a dataset $\mathsf{X} = [x_{i,j}]_{i\in[n],j\in[m]}$, the degree of the ith row is defined as $D(i) \triangleq \sum_{j=1}^{m} x_{i,j}$. The degree-based rank function $R_d : [n] \to [n]$ is characterized by the following relation:

$$\forall i, i' \in [n], i < i' : R_d(i) \leq R_d(i') \iff D(i) \leq D(i').$$

Remark 1. In this work, we have considered a degree-based rank function which does not discriminate between different elements of each row in calculating the degree. The analysis can potentially be extended to weighted-degree-based rank functions, where the degree is computed as a weighted sum of the row elements, i.e, $D(i) \triangleq \sum_{j=1}^{m} w_j x_{i,j}, w_j \in \mathbb{R}, i \in [n]$.

Bob receives the obfuscated dataset Y, and wishes to reconstruct the rank function $R(\cdot)$ associated with X. We consider the conventionally used KRCC (e.g., see [32]) as the distortion criterion measuring the quality of Bob's reconstructed rank function $\widehat{R}(\cdot)$.

Definition 2 (**Kendall Rank Correlation Coefficient**). For two rank functions $R_d(\cdot)$ and $\widehat{R_d}(\cdot)$,

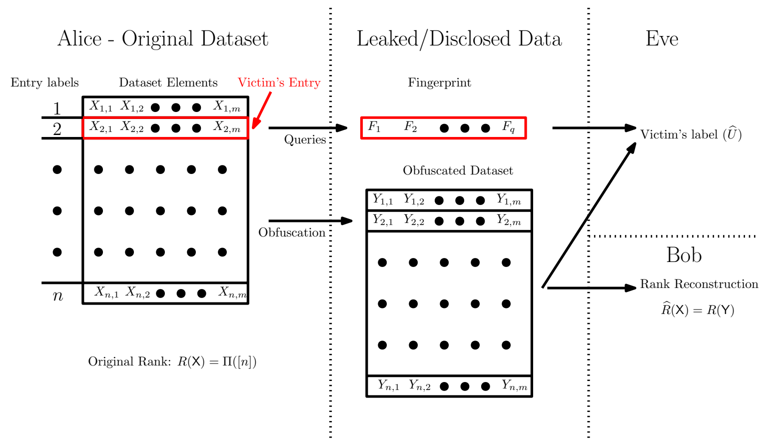

Fig. 2. The dataset obfuscation setup.

their KRCC distance is defined as[1]

$$d_{KRCC}(R_d, \widehat{R}_d) \triangleq \frac{1}{n(n-1)} \sum_{(i,j)\in[n]} \mathbb{1}\Big(R_d(i) > R_d(j) \;\&\; \widehat{R}_d(i) < \widehat{R}_d(j)\Big) \tag{1}$$

k-Letter Obfuscation Strategy: As mentioned in the introduction, a widely used obfuscation method is to perturb each dataset entry independently by passing them through identical test-channels [5], [13]–[15]. We call such mechanisms single-letter obfuscation mechanisms. One justification for their use is that in applications such as search engines and recommendation systems, standard ranking algorithms such as PageRank [29] require both an accurate estimation of the degree-based rank function and a small ℓ_1 distance between the original dataset and the obfuscated dataset for reliable performance. Single-letter obfuscation mechanisms facilitate analyzing and controlling the ℓ_1 distance between the two datasets by appropriate choice of the underlying obfuscating test-channel. A k-letter obfuscation strategy is a generalization of single-letter strategies, where randomly partitioned subsets of size k of elements of each entry are passed through k-letter test-channels for obfuscation. The following formally defines a k-letter obfuscation strategy.

Definition 3 (k-letter Obfuscation Strategy). For a random dataset $\mathsf{X} = [X_{i,j}]_{i\in[n],j\in[m]}$ parametrized by $(n, m, \mathcal{X}, P_X)$, a k-letter obfuscation strategy is parametrized by the conditional distribution

[1] In some texts KRCC is defined as $d'_{KRCC}(R_d, \widehat{R}_d) \triangleq \frac{2}{n(n-1)} \sum_{i<j} sign(R_d(i) - R_d(j)) sign(\widehat{R}_d(i) - \widehat{R}_d(j))$. It can be observed that $d'_{KRCC}(\cdot,\cdot) = 1 - \frac{n}{(n-1)} d_{KRCC}(\cdot,\cdot)$. We adapt the formulation in Definition 2 as it allows for more concise arguments.

$P_{Y^k|X^k}$. The obfuscated dataset $\mathsf{Y} = [Y_{i,j}]_{i\in[n],j\in[m]}$ is produced as follows:[2]

$$P_{\mathsf{Y}|\mathsf{X}}(\mathsf{y}|\mathsf{x}) = \prod_{i\in[n]} \prod_{\ell\in[\frac{m}{k}]} P_{Y^k|X^k}((y_{i,j})_{j\in\mathcal{P}_\ell}|(x_{i,j})_{j\in\mathcal{P}_\ell}),$$

where $\mathsf{x} = [x_{i,j}]_{i\in[n],j\in[m]}$, $\mathsf{y} = [y_{i,j}]_{i\in[n],j\in[m]}$ and $(\mathcal{P}_\ell)_{\ell\in[\frac{m}{k}]}$ is a randomly and uniformly chosen partition of $[m]$ into $\frac{m}{k}$ subsets of size equal to k.

The dataset obfuscation problem is formally defined in the following.

Definition 4 (**k-Letter Dataset Obfuscation Problem**)**.** Given a random dataset parametrized by $(n, m, \mathcal{X}, P_X)$, fingerprint length $q \in \mathbb{N}$, query noise distribution $P_{F|X}$, and $\epsilon > 0$, the k-letter dataset obfuscation problem is to characterize the ϵ-optimal k-letter strategy $P^*_{Y^k|X^k}$, defined as

$$P^*_{Y^k|X^k} \triangleq \mathop{\arg\min}_{P_{Y^k|X^k}: \frac{1}{q} I(U;F^q,\mathsf{Y})<\epsilon} \mathbb{E}(d_{KRCC}(R_d, \widehat{R}_d)),$$

where R and $\widehat{R}$ are the degree-based rank functions associated with X and Y, respectively, and U is uniformly distributed over $[n]$. The set of all pairs (ϵ, δ) for which there exists $P_{Y^k|X^k}$ such that $\mathbb{E}(d_{KRCC}(R, \widehat{R})) < \delta$ and $\frac{1}{q} I(U; F^q, \mathsf{Y}) < \epsilon$ is called the *achievable privacy-utility set* and is denoted by $\mathcal{R}(k, n, m, q, \mathcal{X}, P_X, P_{F|X})$.

In the rest of the paper, for brevity, we denote the achievable privacy-utility region by $\mathcal{R}(n, m, P_X, P_{F|X})$ when the values of k, q and $\mathcal{X}$ are clear from the context.

III. Characterizing the Privacy-Utility Tradeoff

In this section, we consider single-letter obfuscation mechanisms and evaluate their fundamental performance limits, in terms of the utility-privacy tradeoff measured with respect to KRCC utility metric and information leakage privacy metric described in the previous section. The analysis can also be extended to finite-letter obfuscation mechanism using similar techniques. For ease of explanation, the main theorems are provided for binary alphabet datasets.

Recall that given a dataset X, parametrized by $(n, m, \mathcal{X}, P_X)$, and a conditional distribution $P_{Y|X}$, a single-letter obfuscation mechanism produces the obfuscated dataset Y conditioned on

[2]For ease of notation, we have assumed that m is divisible by k.

X by passing each element of X through independent and statistically identical test-channels characterized by $P_{Y|X}$. In order to provide our main results, we first introduce the following notation. Given joint distribution $P_{X,Y} = P_X P_{Y|X}$ on pairs of binary variables (X, Y), we define $Q(P_X, P_{Y|X}) \triangleq F_{N_1,N_2}(0, 0)$, where $F_{N_1,N_2}(\cdot, \cdot)$ is the cumulative distribution function (CDF) of zero-mean jointly Gaussian variables N_1 and N_2 with covariance matrix $\Sigma \triangleq [\sigma_{i,j}]_{i,j\in\{1,2\}}$ given by

$$\sigma_{1,1} \triangleq 2P_X(0)P_X(1), \qquad \sigma_{2,2} \triangleq 2P_Y(0)P_Y(1), \qquad \sigma_{2,1} = \sigma_{1,2} \triangleq 2P_X(1)(P_Y(1) - P_{Y|X}(1|1)). \quad (2)$$

The following provides one of the main results of the paper.

Theorem 1. Let $q, n, m \in \mathbb{N}, \mathcal{X} = \mathcal{F} = \{0, 1\}$, P_X be a probability measure on $\mathcal{X}$, and $P_{F|X}$ a collection of probability measures on $\mathcal{F}$. Then, there exists $b > 0$ such that:

$$\mathcal{R}_{in}(n, m, P_X, P_{F|X}) \subseteq \mathcal{R}(n, m, P_X, P_{F|X}) \subseteq \mathcal{R}_{out}(n, m, P_X, P_{F|X}),$$

where

$$\mathcal{R}_{in}(n, m, P_X, P_{F|X}) \triangleq \bigcup_{P_{Y|X}} \Big\{(\epsilon, \delta) | \epsilon \geq I(P_Y; P_{Y|F}) + \zeta + b\frac{\log^{\frac{3}{2}} m}{\sqrt{m}}, \delta \geq Q(P_X, P_{Y|X}) + \frac{(42\sqrt[4]{2} + 16)}{\sqrt{m}}\theta\gamma\Big\},$$
$$\mathcal{R}_{out}(n, m, P_X, P_{F|X}) \triangleq \bigcup_{P_{Y|X}} \Big\{(\epsilon, \delta) | \epsilon \geq I(P_Y; P_{Y|F}), \delta \geq Q(P_X, P_{Y|X}) - \frac{(42\sqrt[4]{2} + 16)}{\sqrt{m}}\theta\gamma\Big\},$$
$$\theta \triangleq \frac{4}{\sqrt{\lambda^*}}, \qquad \lambda^* \triangleq \min\{\sigma_{1,1} - |\sigma_{1,2}|, \sigma_{2,2} - |\sigma_{2,1}|\},$$
$$\gamma \triangleq 2P(X = Y)P(X \neq Y) + 2^{\frac{5}{2}}(P_{X,Y}(0, 0)P_{X,Y}(1, 1) + P_{X,Y}(0, 1)P_{X,Y}(1, 0)),$$
$$\zeta \triangleq \max(\max_{P_Y}(I(P_Y; P_{Y|F}) - \frac{\log n}{q}), 0),$$

the mutual information $I(P_Y; P_{Y|F})$ is evaluated with respect to $P_{Y,F}$ induced by the Markov chain $Y \leftrightarrow X \leftrightarrow F$, and the union is over all probability distributions $P_{Y|X}$. Particularly, for asymptotically large datasets, we have:

$$\lim_{m\to\infty} \mathcal{R}(n, m, P_X, P_{F|X}) = \bigcup_{P_{Y|X}} \{(\epsilon, \delta) | \epsilon \geq I(P_Y; P_{Y|F}), \delta \geq Q(P_X, P_{Y|X})\}.$$

Proof. Please refer to Appendix A. □

Theorem 1 provides the achievable utility-privacy region as a union of achievable regions for each obfuscating test-channel $P_{Y|X}$. A relevant problem of interest is to find the optimal test channel $P^{\epsilon}_{Y|X}$ minimizing the utility cost δ given a privacy cost ϵ. The following theorem provides a characterization of $P^{\epsilon}_{Y|X}$ in the form of a computable convex optimization problem for asymptotically large datasets, i.e., for $m \to \infty$.

Theorem 2. Let $q, n, m \in \mathbb{N}, \mathcal{X} = \mathcal{F} = \{0, 1\}$, P_X be a probability measure on $\mathcal{X}$, and $P_{F|X}$ be a collection of probability measures on $\mathcal{F}$, such that $\max_{P_Y} I(P_Y; P_{Y|F}) \leq \frac{\log n}{q}$. Define:

$$P^{\epsilon}_{Y|X} \triangleq \operatorname*{arg\,min}_{P_{Y|X}:I(P_Y;P_{Y|F})\leq\epsilon} \{\delta|(\epsilon,\delta) \in \lim_{m\to\infty} \mathcal{R}(n, m, P_X, P_{F|X})\}, \epsilon > 0.$$

Then,

$$P^{\epsilon}_{Y|X} = \operatorname*{arg\,min}_{P_{Y|X}:I(P_Y,P_{Y|F})=\epsilon} \frac{\mathrm{Cov}(N_1, N_2)}{\sqrt{\mathrm{Var}(N_1)\mathrm{Var}(N_2)}}, \tag{3}$$

where N_1 and N_2 are zero-mean jointly Gaussian variables with covariance matrix given in Equation (2).

Proof. Please refer to Appendix B. □

The following follows from the proof of Theorem 2.

Corollary 1. The optimal obfuscating test-channel in Equation (3) can be computed through the following optimization:

$$P^{\epsilon}_{Y|X} = \operatorname*{arg\,max}_{\substack{p_1,p_2:I(P_Y,P_{Y|F})=\epsilon \\ p_1+p_2\leq 1}} \frac{(p_1 + p_2 - 1)^2}{(\overline{p}_1 P_X(0)+p_2 P_X(1))(p_1 P_X(0)+\overline{p}_2 P_X(1))}, \tag{4}$$

where $\overline{p}_i \triangleq 1 - p_i, i = 1, 2$, and $P_{Y|X}(1|0) = p_1, P_{Y|X}(0|1) = p_2$.

We show that the objective function in the optimization in Equation (4) is convex. To see this, let us define $a \triangleq p_1 + p_2$ and $b \triangleq 2P_X(0)p_1 - 2P_X(1)p_2 + 1 - 2P_X(0)$. Then, the objective function can be written as

$$g(a,b) \triangleq \frac{4(a-1)^2}{(1-b)(1+b)} = \frac{4(a-1)^2}{1-b^2}, a \in [0,1], b \in [-1,1].$$

The Hessian matrix of second partial derivatives of $\frac{1}{4}g(a,b)$ is given as:

$$\mathcal{H} = \begin{bmatrix} \frac{\partial^2}{\partial a^2}\frac{1}{4}g(a,b) & \frac{\partial^2}{\partial a \partial b}\frac{1}{4}g(a,b) \\ \frac{\partial^2}{\partial a \partial b}\frac{1}{4}g(a,b) & \frac{\partial^2}{\partial b^2}\frac{1}{4}g(a,b) \end{bmatrix} = \begin{bmatrix} \frac{2}{1-b^2} & \frac{4(a-1)b}{(1-b^2)^2} \\ \frac{4(a-1)b}{(1-b^2)^2} & \frac{2(a-1)^2(1+3b^2)}{(1-b^2)^3} \end{bmatrix}$$

We have:

$$det(\mathcal{H}) = \frac{4(a-1)^2(1-b^2)}{(1-b^2)^4} \geq 0.$$

So, $g(a,b)$ is a convex function, and since (a,b) are a linear transformation of (p_1, p_2), the objective function in Equation (4) is convex in (p_1, p_2). This optimization problem can in general be solved efficiently using numerical methods. There are special cases where exact analytical solution can be derived. For instance, the following corollary characterizes the optimal test-channel if the query noise $P_{F|X}$ is a binary symmetric channel with transition probability q (BSC(q)), and the choice of the obfuscating test-channel is restricted to BSC test-channels, i.e. $P_{Y|X}(0|1) = P_{X|Y}(1|0) = p, p \in [0,1]$.

Corollary 2. Assume that $P_{F|Y}$ is a BSC(q) channel, where $q \in [0, \frac{1}{2}]$, and the choice of obfuscating test-channel $P_{Y|X}$ is restricted to BSC test channels. Then, given $\epsilon > 0$ the optimizing obfuscating test-channel, minimizing the KRCC cost is the BSC parametrized by

$$p = \frac{h_b^{-1}(1-\epsilon) - q}{1-2q},$$

where $h_b^{-1}(\cdot)$ is the inverse of the binary entropy function defined as $h_b(x) = -x\log_2(x) - (1-x)\log_2(1-x), x \in [0, \frac{1}{2}]$.

IV. Numerical Simulations

This section provides numerical simulations to illustrate some of the theoretical derivations provided in the prior sections.

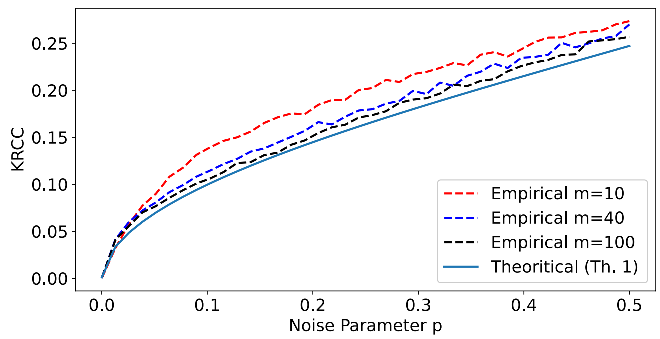

Fig. 3. Comparison of analytical derivation of KRCC with empirical observations through numerical simulation.

A. Analytical and Empirical Simulation of KRCC

In the proof of Theorem 1, we show that the resulting KRCC from obfuscating a dataset X parametrized by (n, m, X, P_X) using a single-letter obfuscation mechanism $P_{Y|X}$ is given by $Q(P_X, P_{Y|X})$. To verify this, we have simulated the obfuscation mechanism when the obfuscating test-channel is $BSC(p), p \in [0, \frac{1}{2}]$ is applied to a dataset with $P_X(0) = P_X(1) = \frac{1}{2}$, $n = 200$, and $m \in \{10, 40, 100\}$. To ensure accuracy, we have performed numerical simulations for each value of m by generating the dataset 40 times, performing obfuscation and calculating the resulting KRCC. Figure 3 shows the resulting analytical and empirically observed KRCCs. As can be observed the analytical result is close to the empirical performance and the empirical KRCC converges to the analytical derivation as m becomes larger.

B. Asymmetric Obfuscating Test-Channels

In Section III, we argue that the objective function of Equation (4) is convex which implies the KRCC is concave as a function of $p_1 = P_{Y|X}(1|0)$ and $p_2 = P_{Y|X}(0|1)$. Figure 4 shows the KRCC when $P_X(0) = P_X(1) = \frac{1}{2}$ and $m \to \infty$ for $p_1, p_2 \in [0, \frac{1}{2}]$. It can be observed that KRCC is convex in (p_1, p_2) as predicted.

C. Privacy-Utility Tradeoff

In Figure 5, we have shown the privacy-utility tradeoff for the scenario where a symmetric dataset ($P_X(1) = \frac{1}{2}$) is obfuscated using an optimal symmetric test-channel ($P_{Y|X}(0|1) = P_{Y|X}(1|0)$),

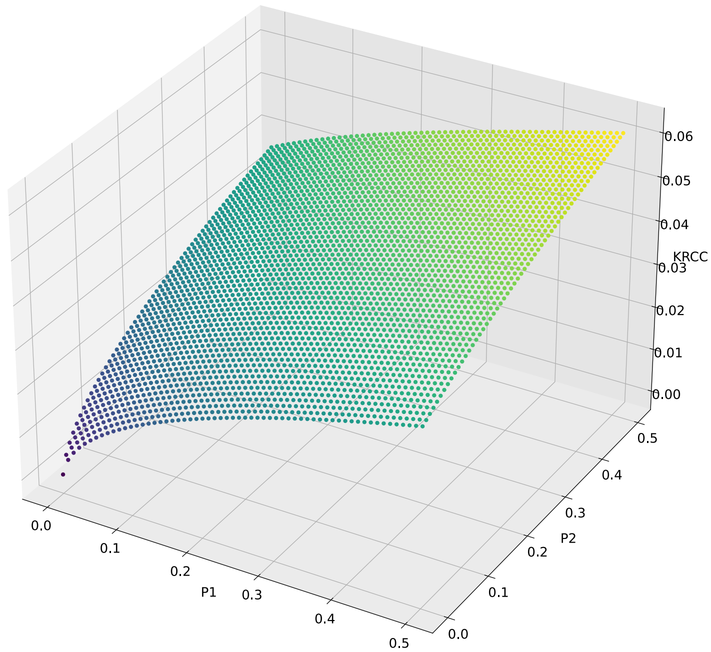

Fig. 4. KRCC values for asymmetric obfuscation mechanisms.

and the query noise is modeled by a BSC(0.1). The resulting achievable privacy-utility region $\mathcal{R}(n, m, P_X, P_{F|X})$ is shown as the blue shaded region in the figure. The optimal symmetric test-channel used in the simulation is derived using Corollary 2 in the previous section.

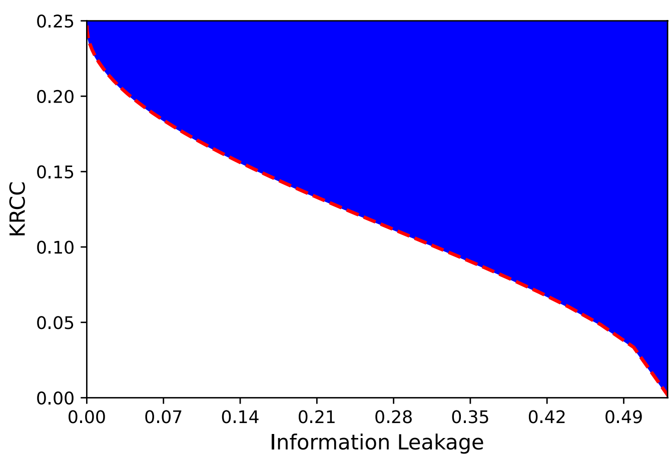

Fig. 5. Privacy-Utility Tradeoff. The shaded region indicates the achievable privacy-utility set.

V. Conclusion

We have considered the privacy-utility tradeoff in dataset obfuscation, where utility is measured with respect to KRCC metric and privacy is measured as privacy leakage under fingerprinting attacks. We have quantified a fundamental trade-off between rank-preservation and user anonymity. We have considered single-letter obfuscation mechanisms and their fundamental performance limits were characterized. We have characterized the optimal obfuscating test-channel, optimizing the privacy-utility tradeoff in the form of a convex optimization problem which can be solved efficiently.

References

[1] Hao Wang, Lisa Vo, Flavio P Calmon, Muriel Médard, Ken R Duffy, and Mayank Varia. Privacy with estimation guarantees. *IEEE Transactions on Information Theory*, 65(12):8025–8042, 2019.

[2] David Salomon. *Data privacy and security: encryption and information hiding*. Springer Science & Business Media, 2003.

[3] Kelly D Martin and Patrick E Murphy. The role of data privacy in marketing. *Journal of the Academy of Marketing Science*, 45(2):135–155, 2017.

[4] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *2009 30th IEEE symposium on security and privacy*, pages 173–187. IEEE, 2009.

[5] Nazanin Takbiri, Amir Houmansadr, Dennis L Goeckel, and Hossein Pishro-Nik. Matching anonymized and obfuscated time series to users' profiles. *IEEE Transactions on Information Theory*, 65(2):724–741, 2018.

[6] Joseph A Calandrino, Ann Kilzer, Arvind Narayanan, Edward W Felten, and Vitaly Shmatikov. " you might also like:" privacy risks of collaborative filtering. In *2011 IEEE symposium on security and privacy*, pages 231–246. IEEE, 2011.

[7] Mahshad Shariatnasab, Farhad Shirani, and Elza Erkip. Fundamental privacy limits in bipartite networks under active attacks. *IEEE Journal on Selected Areas in Communications*, 40(3):940–954, 2022.

[8] Gilbert Wondracek, Thorsten Holz, Engin Kirda, and Christopher Kruegel. A practical attack to de-anonymize social network users. In *2010 ieee symposium on security and privacy*, pages 223–238. IEEE, 2010.

[9] Minoh Jeong, Alex Dytso, and Martina Cardone. Ranking recovery under privacy considerations. *Transactions on Machine Learning Research*, 2022.

[10] Daniel Alabi, Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. Private rank aggregation in central and local models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5984–5991, 2022.

[11] Ziqi Yan, Gang Li, and Jiqiang Liu. Private rank aggregation under local differential privacy. *International Journal of Intelligent Systems*, 35(10):1492–1519, 2020.

[12] Shang Shang, Tiance Wang, Paul Cuff, and Sanjeev Kulkarni. The application of differential privacy for rank aggregation: Privacy and accuracy. In *17th International Conference on Information Fusion (FUSION)*, pages 1–7. IEEE, 2014.

[13] Runting Shi, Richard Chow, and Tsz Hong Hubert Chan. Privacy-preserving aggregation of time-series data, 2016.

[14] Mayra Zurbarán, Karen Avila, Pedro Wightman, and Michael Fernandez. Near-rand: Noise-based location obfuscation based on random neighboring points. *IEEE Latin America Transactions*, 13(11):3661–3667, 2015.

[15] Pedro Wightman, Winston Coronell, Daladier Jabba, Miguel Jimeno, and Miguel Labrador. Evaluation of location obfuscation techniques for privacy in location based information systems. In *2011 IEEE Third Latin-American Conference on Communications*, pages 1–6. IEEE, 2011.

[16] Amirreza Zamani, Tobias J Oechtering, and Mikael Skoglund. Data disclosure with non-zero leakage and non-invertible leakage matrix. *IEEE Transactions on Information Forensics and Security*, 17:165–179, 2021.

[17] Borzoo Rassouli and Deniz Gündüz. Optimal utility-privacy trade-off with total variation distance as a privacy measure. *IEEE Transactions on Information Forensics and Security*, 15:594–603, 2019.

[18] Lalitha Sankar, S Raj Rajagopalan, and H Vincent Poor. Utility-privacy tradeoffs in databases: An information-theoretic approach. *IEEE Transactions on Information Forensics and Security*, 8(6):838–852, 2013.

[19] Jiachun Liao, Oliver Kosut, Lalitha Sankar, and Flavio du Pin Calmon. Tunable measures for information leakage and applications to privacy-utility tradeoffs. *IEEE Transactions on Information Theory*, 65(12):8043–8066, 2019.

[20] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622, 2001.

[21] Michael Hay, Liudmila Elagina, and Gerome Miklau. Differentially private rank aggregation. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 669–677. SIAM, 2017.

[22] JiaQi Liu, XueRong Li, and JiChang Dong. A survey on network node ranking algorithms: Representative methods, extensions, and applications. *Science China Technological Sciences*, 64(3):451–461, 2021.

[23] Bryan Brancotte, Bo Yang, Guillaume Blin, Sarah Cohen-Boulakia, Alain Denise, and Sylvie Hamel. Rank aggregation with ties: Experiments and analysis. *Proceedings of the VLDB Endowment (PVLDB)*, 8(11):1202–1213, 2015.

[24] N. Takbiri, R. Soltani, D.L. Goeckel, A. Houmansadr, and H. Pishro-Nik. Asymptotic loss in privacy due to dependency in Gaussian traces. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2019.

[25] Jessica Su, Ansh Shukla, Sharad Goel, and Arvind Narayanan. De-anonymizing web browsing data with social networks. In *Proceedings of the 26th international conference on world wide web*, pages 1261–1269, 2017.

[26] J. Domingo-Ferrer, D. Sánchez, and J. Soria-Comas. Database anonymization: privacy models, data utility, and microaggregation-based inter-model connections. *Synthesis Lectures on Information Security, Privacy, & Trust*, 8(1):1–136, 2016.

[27] M. Smith, C. Disselkoen, S. Narayan, F. Brown, and D. Stefan. Browser history re: visited. In *12th {USENIX} Workshop on Offensive Technologies ({WOOT} 18)*, 2018.

[28] Konstantinos Solomos, John Kristoff, Chris Kanich, and Jason Polakis. Tales of favicons and caches: Persistent tracking in modern browsers. In *Network and Distributed System Security Symposium*, 2021.

[29] Fan Chung. A brief survey of pagerank algorithms. *IEEE Trans. Netw. Sci. Eng.*, 1(1):38–42, 2014.

[30] Pavel Berkhin. A survey on pagerank computing. *Internet mathematics*, 2(1):73–120, 2005.

[31] Elli Voudigari, Nikos Salamanos, Theodore Papageorgiou, and Emmanuel J Yannakoudakis. Rank degree: An efficient algorithm for graph sampling. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 120–129. IEEE, 2016.

[32] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[33] Martin Raič. A multivariate berry–esseen theorem with explicit constants. 2019.

[34] Sadid Sahami, Gene Cheung, and Chia-Wen Lin. Fast graph sampling for short video summarization using gershgorin disc alignment. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1765–1769. IEEE, 2022.

[35] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

[36] Yury Polyanskiy and Sergio Verdú. Empirical distribution of good channel codes with nonvanishing error probability. *IEEE transactions on information theory*, 60(1):5–21, 2013.

Appendix A
Proof of Theorem 1

Consider a fixed $n, m, q \in \mathbb{N}$, a distribution P_X, conditional distributions $P_{F|X}$ and $P_{Y|X}$. We first evaluate the resulting KRCC measure when a single-letter obfuscation mechanism $P_{Y|X}$ is applied to a dataset X parametrized by $(n, m, \mathcal{X}, P_X)$. Let $R_d(\cdot)$ and $\widehat{R}_d(\cdot)$ denote the degree-based rank function associated with the original dataset X and obfuscated dataset Y, respectively. Then,

$$
\begin{aligned}
&\mathbb{E}(d_{KRCC}(R_d, \widehat{R}_d)) \\
&= \mathbb{E}\Big(\frac{1}{n(n-1)} \sum_{k,l=1}^{n} \mathbb{1}\Big(R_d(k) > R_d(l) \& \widehat{R}_d(k) < \widehat{R}_d(l)\Big)\Big) \\
&\overset{(a)}{=} \frac{1}{n(n-1)} \sum_{k,l=1}^{n} P\Big(R_d(k) > R_d(l) \& \widehat{R}_d(k) < \widehat{R}_d(l)\Big) \\
&= \frac{1}{n(n-1)} \sum_{k,l=1}^{n} P\Big(\frac{1}{\sqrt{m}} R_d(k) > \frac{1}{\sqrt{m}} R_d(l) \& \frac{1}{\sqrt{m}} \widehat{R}_d(k) < \frac{1}{\sqrt{m}} \widehat{R}_d(l)\Big) \\
&\overset{(b)}{=} \frac{1}{n(n-1)} \sum_{k,l=1}^{n} P\Big(\frac{1}{\sqrt{m}} D_{\mathsf{X}}(k) > \frac{1}{\sqrt{m}} D_{\mathsf{X}}(l) \ \& \ \frac{1}{\sqrt{m}} D_{\mathsf{Y}}(k) < \frac{1}{\sqrt{m}} D_{\mathsf{Y}}(l)\Big) \\
&\overset{(c)}{=} P\Big(\frac{1}{\sqrt{m}} D_{\mathsf{X}}(1) > \frac{1}{\sqrt{m}} D_{\mathsf{X}}(2) \& \frac{1}{\sqrt{m}} D_{\mathsf{Y}}(1) < \frac{1}{\sqrt{m}} D_{\mathsf{Y}}(2)\Big) \\
&= P\Big(\frac{1}{\sqrt{m}} \sum_{j=1}^{m} X_{1,j} > \frac{1}{\sqrt{m}} \sum_{j=1}^{m} X_{2,j} \& \\
&\qquad\qquad\qquad \frac{1}{\sqrt{m}} \sum_{j=1}^{m} Y_{1,j} < \frac{1}{\sqrt{m}} \sum_{j=1}^{m} Y_{2,j}\Big) \\
&= P\Big(\frac{1}{\sqrt{m}} \sum_{j=1}^{m} X_{2,j} - X_{1,j} < 0 \& \frac{1}{\sqrt{m}} \sum_{j=1}^{m} Y_{1,j} - Y_{2,j} < 0\Big),
\end{aligned} \tag{5}
$$

where (a) follows the form linearity of expectation, (b) uses the definition of degree-based rank function (Definition 1), and (c) follows from the fact that the original dataset elements are IID and in single-letter obfuscation mechanisms the obfuscating test-channels are statistically identical. We bound the last term using a generalization of the Berry-Esseen result to multivariate scenarios given in [33, Theorem 1.1]. The theorem is stated below for completeness.

Theorem 3. [[33], Theorem 1.1] Let $(Z_{1,i}, Z_{2,i}), i \in [m]$ be independent pairs of sequences of independent, zero-mean, and unit-variance random variables, where $m \in \mathbb{N}$, and let $W'_j \triangleq$

$\frac{1}{\sqrt{m}}\sum_{i\in[m]} Z_{j,i}$, $j \in \{1,2\}$. Then, for any measurable convex set $\mathcal{A}$,

$$|P((W_1', W_2')\in\mathcal{A}) - P((N_1', N_2') \in \mathcal{A})| \leq \frac{(42\sqrt[4]{2} + 16)}{m\sqrt{m}} \sum_{i=1}^{m} \mathbb{E}|\underline{Z}_i|^3,$$

where (N_1', N_2') is a pair of independent and unit-variance Gaussian random variables, and $|\underline{Z}_i| \triangleq \sqrt{Z_{1,i}^2 + Z_{2,i}^2}, i \in [m]$.

We let $W_1 \triangleq \frac{1}{\sqrt{m}} \sum_{j=1}^{m} X_{2,j} - X_{1,j}$ and $W_2 \triangleq \frac{1}{\sqrt{m}} \sum_{j=1}^{m} Y_{1,j} - Y_{2,j}$. Then, W_1 and W_2 are zero-mean variables since $X_{i,j}, i \in [n], j \in [m]$ are IID and the obfuscating test-channels are statistically identical so that $Y_{i,j}, i \in [n], j \in [m]$ are IID. Next, we find the covariance matrix of (W_1, W_2). First, we find the variance of W_1:

$$\begin{aligned}
&Var(W_1) = \mathbb{E}(W_1^2) = \mathbb{E}\left(\left(\frac{1}{\sqrt{m}}\sum_{j=1}^{m}(X_{1,j} - X_{2,j})\right)^2\right)\\
&\overset{(a)}{=} \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\Big(\big(X_{2,j} - X_{1,j}\big)^2\Big) + \frac{1}{m}\sum_{i\neq j}\mathbb{E}\Big(\big(X_{1,i} - X_{2,i})(X_{1,j} - X_{2,j}\big)\Big)\\
&\overset{(b)}{=} \mathbb{E}\Big((X_{1,1} - X_{2,1})^2\Big) \overset{(c)}{=} P(X_{1,1} \neq X_{2,1}) = P_{X_1,X_2}(0,1) + P_{X_1,X_2}(1,0) \overset{(d)}{=} 2P_X(0)P_X(1),
\end{aligned}$$

where in (a) we have used the fact that the dataset elements are identically distributed and the test-channels are statistically identical, (b) follows since $(X_{1,i}, X_{2,i})$ is independent of $(X_{1,j}, X_{2,j})$ since the dataset elements are IID, (c) follows since $X_{1,1} - X_{2,1} \in \{-1, 0, 1\}$, and (d) follows since X_1 and X_2 are IID. Variance of W_2 is similarly derived as:

$$Var(W_2) = 2P_Y(0)P_Y(1) = 2\Big(P_X(0)P_{Y|X}(0|0) + P_X(1)P_{Y|X}(0|1)\Big)\Big(P_X(0)P_{Y|X}(1|0) + P_X(1)P_{Y|X}(1|1)\Big)$$

The covariance between W_1 and W_2 is given by:

$$\begin{aligned}
&Cov(W_1, W_2) = \mathbb{E}(W_1W_2) = \mathbb{E}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}(X_{2,i} - X_{1,i}))\frac{1}{\sqrt{m}}\sum_{j=1}^{m}(Y_{1,j} - Y_{2,j}))\right)\\
&= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}((X_{2,i} - X_{1,i})(Y_{1,i} - Y_{2,i})) = \mathbb{E}((X_{2,1} - X_{1,1})(Y_{1,1} - Y_{2,1})) = 2\mathbb{E}(X_{2,1}Y_{1,2}) - 2\mathbb{E}(X_{1,1}Y_{1,1})\\
&= 2P_X(1)P_Y(1) - 2P_{X,Y}(1,1) = 2P_X(1)(P_Y(1) - P_{Y|X}(1,1))
\end{aligned}$$

Let Σ_{W_1,W_2} be the covariance matrix of (W_1, W_2). We define $\underline{W}$ as the column vector consisting of W_1, W_2 and define $\underline{W'} \triangleq \Sigma^{\frac{-1}{2}}_{W_1,W_2}\underline{W}$. Then,

$$W'_1 = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \Sigma^{\frac{-1}{2}}_{W_1,W_2}(1,1)(X_{2,j} - X_{1,j}) + \Sigma^{\frac{-1}{2}}_{W_1,W_2}(1,2)(Y_{1,j} - X_{2,j}),$$
$$W'_2 = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \Sigma^{\frac{-1}{2}}_{W_1,W_2}(2,1)(X_{2,j} - X_{1,j}) + \Sigma^{\frac{-1}{2}}_{W_1,W_2}(2,2)(Y_{1,j} - X_{2,j}),$$

where $\Sigma^{-\frac{1}{2}}_{W_1,W_2}(i,j), i,j \in \{1,2\}$ is the (i,j)th element of the matrix $\Sigma^{-\frac{1}{2}}_{W_1,W_2}$. It should be noted that $\Sigma^{\frac{-1}{2}}_{W_1,W_2}$ exists since Σ_{W_1,W_2} is positive semi-definite. It is straightforward to check that W'_1, W'_2 are zero-mean and unit variance. Consequently, W'_1, W'_2 satisfy the properties of Theorem 3. Let

$$\mathcal{A} \triangleq \{\underline{w'} \in \mathbb{R}^2 : \Sigma^{\frac{1}{2}}_{W_1,W_2}\underline{w'} \leq 0\}$$

Then, by Theorem 3, we have:

$$|P((W'_1, W'_2) \in \mathcal{A}) - P((N'_1, N'_2) \in \mathcal{A})| \leq \frac{(42\sqrt[4]{2} + 16)}{\sqrt{m}} \mathbb{E}|\underline{Z}|^3,$$

where $\underline{Z} = (Z_1, Z_2)$, and

$$Z_1 = \Sigma^{\frac{-1}{2}}_{W_1,W_2}(1,1)(X_{2,1} - X_{1,1}) + \Sigma^{\frac{-1}{2}}_{W_1,W_2}(1,2)(Y_{1,1} - X_{2,1})$$
$$Z_2 = \Sigma^{\frac{-1}{2}}_{W_1,W_2}(2,1)(X_{2,1} - X_{1,1}) + \Sigma^{\frac{-1}{2}}_{W_1,W_2}(2,2)(Y_{1,1} - X_{2,1}).$$

Note that

$$|\underline{Z}| \leq |\Sigma^{-\frac{1}{2}}_{W_1,W_2}|_F |[X_{2,1} - X_{1,1}, Y_{1,2} - Y_{2,1}]|,$$

where $|\cdot|_F$ is the Frobenius norm. Let $\Sigma_{W_1,W_2} = V\Lambda V^*$, where V and Λ are the singular value decomposition matrices associated with Σ_{W_1,W_2}, V is unitary, and V^* is the conjugate transpose of V. So, $\Sigma^{-\frac{1}{2}}_{W_1,W_2} = \Lambda^{-\frac{1}{2}}V^*$. So, $|\Sigma^{-\frac{1}{2}}_{W_1,W_2}|_F \leq |\Lambda^{-\frac{1}{2}}|_F |V^*|_F \leq \frac{4}{\sqrt{\lambda^*}}$, where λ^* is the smallest eigenvalue of Σ_{W_1,W_2} and we have used the fact that $|V^*|_F = trace(VV^*) = trace(I_2) = 2$. Furthermore, by the Gershgorin circle theorem [34], [35], we have $\lambda^* \geq \max(\sigma_{1,1} - |\sigma_{1,2}|, \sigma_{2,2} - |\sigma_{2,1}|)$. So,

$|\Sigma_{W_1,W_2}^{-\frac{1}{2}}|_F \leq \frac{4}{\lambda^*} = \theta$. Let $\gamma \triangleq \mathbb{E}(|[X_{2,1} - X_{1,1}, Y_{1,2} - Y_{2,1}]|)$. Then, γ is given by:

$$\begin{aligned}
\gamma &= \mathbb{E}\left(\left((X_{2,1} - X_{1,1})^2 + (Y_{1,1} - Y_{2,1})^2\right)^{\frac{3}{2}}\right) \\
&= P(|X_{2,1} - X_{1,1}| = 0, |Y_{1,1} - Y_{2,1}| = 1) \\
&+ P(|X_{2,1} - X_{1,1}| = 1, |Y_{1,1} - Y_{2,1}| = 0) \\
&+ 2^{\frac{3}{2}} P(|X_{2,1} - X_{1,1}| = 1, |Y_{1,1} - Y_{2,1}| = 1) \\
&= 2P(X = Y)P(X \neq Y) + 2^{\frac{5}{2}}(P_{X,Y}(0,0)P_{X,Y}(1,1) + P_{X,Y}(0,1)P_{X,Y}(1,0))
\end{aligned}$$

So far, we have shown that:

$$|P((W_1', W_2') \in \mathcal{A}) - P((N_1', N_2') \in \mathcal{A})| \leq \frac{(42\sqrt[4]{2} + 16)}{\sqrt{m}} \theta\gamma.$$

Let $\underline{N} = [N_1, N_2]$, where $\underline{N} = \Sigma_{W_1,W_2}^{\frac{1}{2}} \underline{N'}$. It is straightforward to show that $P((N_1', N_2') \in \mathcal{A}) = F_{N_1,N_2}(0,0) = Q(P_X, P_{Y|X})$ and that $\Sigma_{N_1,N_2} = \Sigma_{W_1,W_2}$. As a result, from Equation (5), we have:

$$|\mathbb{E}(d_{KRCC}(R_d, \widehat{R}_d)) - Q(P_X, P_{Y|X})| \leq \frac{(42\sqrt[4]{2} + 16)}{\sqrt{m}} \theta\gamma.$$

Next, we evaluate the privacy cost. We have:

$$\begin{aligned}
I(U; F^q, \mathsf{Y}) &= I(U; \mathsf{Y}) + I(U; F^q|\mathsf{Y}) \\
&\overset{(a)}{=} I(U; F^q|\mathsf{Y}) = I(U, \mathsf{Y}; F^q) - I(\mathsf{Y}; F^q),
\end{aligned}$$

where we have used the chain rule of mutual information in the first and last equality, and in (a) we have used the fact that the dataset is independent of the identity of the victim (recall that the victim is chosen randomly and uniformly from the dataset members, independently of dataset elements). Furthermore, we have:

$$\frac{1}{q} I(U, \mathsf{Y}; F^q) = \frac{1}{q} \sum_{j=1}^{q} I(Y_{U,i_j}; F_j) = I(P_Y, P_{F|Y}).$$

Additionally,

$$I(\mathsf{Y};F^q)=D_{KL}(P_{F^q}||P_{F^q|\mathsf{Y}})$$

Note that Y is a random unstructured code with single-letter distribution P_Y and hence is a *good code* for a channel with transition probability $P_{Y|F}$ and from [36, Theorem 7], we have

$$D_{KL}(P_{F^q}||P_{F^q|\mathsf{Y}})\leq\zeta+b\frac{\log^{\frac{3}{2}}m}{\sqrt{m}}.$$

for some $b>0$. Consequently,

$$I(P_Y,P_{F|Y})\leq\frac{1}{q}I(U;F^q,\mathsf{Y})\leq I(P_Y,P_{F|Y})+\zeta+b\frac{\log^{\frac{3}{2}}m}{\sqrt{m}}.$$

This completes the proof. □

Appendix B

Proof of Theorem 2

From Theorem 1, we need to solve the following optimization problem:

$$\begin{aligned}P^{\epsilon}_{Y|X}&=\underset{P_{Y|X}:I(U;F^q,\mathsf{Y})<\epsilon}{\arg\min}F_{N_1,N_2}(0,0)\\&=\underset{P_{Y|X}:I(U;F^q,\mathsf{Y})<\epsilon}{\arg\min}\int_{n_1=-\infty}^{0}\int_{n_2=-\infty}^{0}\frac{1}{2\pi\sqrt{|\Sigma|}}\exp\Big(-\frac{1}{2}\underline{n}^t\Sigma^{-1}\underline{n}\Big)dn_1dn_2\end{aligned}$$

where, $\underline{n}=\begin{bmatrix}n_1\\n_2\end{bmatrix}$.

Let us define $p_1\triangleq P_{Y|X}(1|0),p_2\triangleq P_{Y|X}(0|1)$. and define variables $N'_1=\frac{N_1}{\sqrt{2Var(N_1)}},N'_2=\frac{N_2}{\sqrt{2Var(N_2)}}$. Note that $P(N_1\leq 0,N_2\leq 0)=P(N'_1\leq 0,N'_2\leq 0)$. The covariance of N'_1,N'_2 is:

$$\Sigma'=\begin{bmatrix}\frac{1}{2}&\frac{1}{2}-p'\\\frac{1}{2}-p'&\frac{1}{2}\end{bmatrix}$$

where p' is defined as:

$$p'\triangleq\frac{1}{2}-\frac{\mathrm{Cov}(N_1,N_2)}{2\sqrt{\mathrm{Var}(N_1)\mathrm{Var}(N_2)}}.\tag{6}$$

So,

$$P^{\epsilon}_{Y|X} = \underset{p_1,p_2:I(U;F^q,\mathsf{Y})<\epsilon}{\arg\min} F_{N_1',N_2'}(0,0)$$
$$= \underset{p_1,p_2:I(U;F^q,\mathsf{Y})<\epsilon}{\arg\min} \int_{n_1'=-\infty}^{0}\int_{n_2'=-\infty}^{0} \frac{1}{2\pi\sqrt{|\Sigma'|}} \exp\left(-\frac{1}{2}\underline{n'}^t\Sigma'^{-1}\underline{n'}\right)d\underline{n'}$$

The eigenvalues of the covariance matrix Σ' are $\lambda_1 = 1-p', \lambda_2 = p'$ and the eigenvectors are the columns of $V = \frac{1}{\sqrt{2}}\begin{bmatrix} \frac{1}{\sqrt{1-p'}} & \frac{1}{\sqrt{p'}} \\ \frac{1}{\sqrt{1-p'}} & -\frac{1}{\sqrt{p'}} \end{bmatrix}$. Let us define $\underline{n}'' = V^t\underline{n}'$ Changing the variables in the integral, we have:

$$\underset{p_1,p_2:I(U;F^q,\mathsf{Y})<\epsilon}{\arg\min} \int_{n_1''=-\infty}^{0}\int_{n_2''=\sqrt{\frac{1-p'}{p'}}n_1''}^{-\sqrt{\frac{1-p'}{p'}}n_1''} \frac{1}{2\pi}\exp(-\frac{n_1''^2+n_2''^2}{2})d_{n_1''}d_{n_2''}$$

Form Equation (6) and using Cauchy-Schwarz inequality we have $0<p'<1$, so the inner integral interval $[\sqrt{\frac{1-p'}{p'}}n_1'', -\sqrt{\frac{1-p'}{p'}}n_1'']$ is decreasing as function of p' for all values of $n_1''<0$. Hence, to minimize the utility cost, we should take the maximum value of p' such that $I(U;F^q,\mathsf{Y})<\epsilon$. That is, the optimization becomes

$$\underset{p_1,p_2:I(P_Y,P_{Y|F})\le\epsilon}{\arg\max} p' = \underset{p_1,p_2:I(P_Y,P_{Y|F})\le\epsilon}{\arg\min} \frac{\mathrm{Cov}(N_1,N_2)}{\sqrt{\mathrm{Var}(N_1)\mathrm{Var}(N_2)}}.$$

From the proof of Theorem 1, we have

$$Cov(N_1,N_2) = P_X(1)(P_Y(1)-P_{Y|X}(1|1))$$
$$= P_X(1)(p_1P_X(0)+(1-p_2)P_X(1)-P_{Y|X}(1|1))$$
$$= P_X(1)P_X(0)(p_1+p_2-1).$$

So, the optimization can be re-written as:

$$P^{\epsilon}_{Y|X} = \underset{p_1,p_2:I(P_Y,P_{Y|F})\le\epsilon}{\arg\min} \frac{p_1+p_2-1}{\sqrt{(\overline{p}_1P_X(0)+p_2P_X(1))(p_1P_X(0)+\overline{p}_2P_X(1))}},$$

where $\overline{p}_1 \triangleq 1-p_1$ and $\overline{p}_2 \triangleq 1-p_2$. Note that we can restrict the minimization to $p_1+p_2<1$ since if $p_1+p_2>1$ the objective function in the above optimization is positive and cannot achieve the minimum value since for $p_1 + p_2 = 0$ the value of zero is already achieved. So,

$$P^{\epsilon}_{Y|X} = \underset{\substack{p_1, p_2 : I(P_Y, P_{Y|F}) \leq \epsilon \\ p_1 + p_2 < 1}}{\arg\min} \frac{p_1 + p_2 - 1}{\sqrt{(\overline{p}_1 P_X(0) + p_2 P_X(1))(p_1 P_X(0) + \overline{p}_2 P_X(1))}}.$$

Next, we show that $\frac{\mathrm{Cov}(N_1, N_2)}{\sqrt{\mathrm{Var}(N_1)\mathrm{Var}(N_2)}}$ is i) increasing in p_1 for any fixed p_2, and ii) increasing in p_2 for any fixed p_1. Hence, we conclude that the optimal value is achieved at the boundary where $I(P_Y, P_{Y|F}) = \epsilon$ since for any point not on the boundary, one either reduce p_1 or p_2 without violating $I(P_Y, P_{Y|F}) \leq \epsilon$. To show i) it suffices to show that $\frac{\mathrm{Cov}^2(N_1, N_2)}{\mathrm{Var}(N_1)\mathrm{Var}(N_2)}$ is decreasing in p_1 for any fixed p_2 since $\mathrm{Cov}^2(N_1, N_2) < 0$ for $p_1 + p_2 < 1$. That is, we wish to show that the following function is decreasing in p_1 for fixed p_2:

$$\frac{(p_1 + p_2 - 1)^2}{(\overline{p}_1 P_X(0) + p_2 P_X(1))(p_1 P_X(0) + \overline{p}_2 P_X(1))} = \frac{p_1 + p_2 - 1}{\overline{p}_1 P_X(0) + p_2 P_X(1)} \frac{p_1 + p_2 - 1}{p_1 P_X(0) + \overline{p}_2 P_X(1)}.$$

Taking derivative of each component in the multiplication with respect to p_1, we have:

$$\frac{\partial}{\partial p_1} \frac{p_1 + p_2 - 1}{\overline{p}_1 P_X(0) + p_2 P_X(1)} = \frac{\overline{p}_1 P_X(0) + p_2 P_X(1) + P_X(0)(p_1 + p_2 - 1)}{(\overline{p}_1 P_X(0) + p_2 P_X(1))^2} = \frac{p_2}{(\overline{p}_1 P_X(0) + p_2 P_X(1))^2}$$

and

$$\frac{\partial}{\partial p_1} \frac{p_1 + p_2 - 1}{p_1 P_X(0) + \overline{p}_2 P_X(1)} = \frac{p_1 P_X(0) + \overline{p}_2 P_X(1) - P_X(0)(p_1 + p_2 - 1)}{(p_1 P_X(0) + \overline{p}_2 P_X(1))^2} = \frac{\overline{p}_2}{(\overline{p}_1 P_X(0) + p_2 P_X(1))^2}$$

Since both derivatives are positive, and both functions are negative-valued, the multiplication has a derivative which is negative with respect to p_1 for fixed p_2. So, $\frac{\mathrm{Cov}^2(N_1, N_2)}{\mathrm{Var}(N_1)\mathrm{Var}(N_2)}$ is decreasing in p_1 for any fixed p_2. Hence, $\frac{\mathrm{Cov}(N_1, N_2)}{\sqrt{\mathrm{Var}(N_1)\mathrm{Var}(N_2)}}$ is increasing in p_1 for any fixed p_2 which proves i). The proof of ii) follows similarly. We conclude that the optimum is achieved at $I(P_Y, P_{Y|F}) = \epsilon$. This completes the proof. $\square$