University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Ding, N;Farokhi, F

Title:
Developing Non-Stochastic Privacy-Preserving Policies Using Agglomerative Clustering

Date:
2020-06-15

Citation:
Ding, N. & Farokhi, F. (2020). Developing Non-Stochastic Privacy-Preserving Policies Using Agglomerative Clustering. IEEE Transactions on Information Forensics and Security, 15, pp.3911-3923. https://doi.org/10.1109/TIFS.2020.3002479.

Persistent Link:
https://hdl.handle.net/11343/251364

Developing Non-Stochastic Privacy-Preserving Policies Using Agglomerative Clustering

Ni Ding, *Member, IEEE,* and Farhad Farokhi, *Senior Member, IEEE*

***Abstract*—We consider a non-stochastic privacy-preserving problem in which an adversary aims to infer sensitive information S from publicly accessible data X without using statistics. We consider the problem of generating and releasing a quantization $\hat{X}$ of X to minimize the privacy leakage of S to $\hat{X}$ while maintaining a certain level of utility (or, inversely, the quantization loss). The variables S and X are treated as bounded and non-probabilistic, but are otherwise general. We consider two existing non-stochastic privacy measures, namely the maximum uncertainty reduction $L_0(S \to \hat{X})$ and the refined information $I_*(S;\hat{X})$ (also called the maximin information) of S. For each privacy measure, we propose a corresponding agglomerative clustering algorithm that converges to a locally optimal quantization solution $\hat{X}$ by iteratively merging elements in the alphabet of X. To instantiate the solution to this problem, we consider two specific utility measures, the worst-case resolution of X by observing $\hat{X}$ and the maximal distortion of the released data $\hat{X}$. We show that the value of the maximin information $I_*(S;\hat{X})$ can be determined by dividing the confusability graph into connected subgraphs. Hence, $I_*(S;\hat{X})$ can be reduced by merging nodes connecting subgraphs. The relation to the probabilistic information-theoretic privacy is also studied by noting that the Gács-Körner common information is the stochastic version of I_* and indicates the attainability of statistical indistinguishability.**

***Index Terms*—Privacy, Information Leakage, Non-stochastic Information Theory.**

I. Introduction

NOWADAYS we share and exchange data with others regularly while being increasingly concerned about whether our personal data is well protected. In particular, in this big data era, advances in efficient data analytics have improved an adversary's ability in obtaining individuals' private information [1]. For example, machine learning models with high accuracy can potentially reveal individual labels of the training data as in the case of membership information attacks [2]. In these cases, the privacy attack can happen during the legitimate use of data, e.g., querying the dataset, survey report, and speech recognition, without direct access to confidential data. Thus, the meaning of privacy has moved far beyond its original definition in [3] and cannot be achieved by just 'anonymizing' or 'secluding' the sensitive data from release.

For a user in the public domain who is interested in the aggregated statistics and interacts with a data curator using queries, the differential privacy (DP) [4] provides a mathematical definition for the privacy loss: an upper bound on how the statistics of sanitized/randomized responses to a query change with and without an individual's record. By replacing query response with data mining, e.g., maximum *a posteriori* estimate, DP can be tuned to a privacy measure in statistical inference problems, e.g., in machine learning [5], [6]. While DP is concerned with the individual indistinguishability of the released data, privacy in information theory [7]–[9] refers to the exact amount (e.g., the number of bits) of sensitive information that is leaked to the public. This enables study of privacy-utility tradeoff (PUT) via the problem of minimizing the privacy leakage subject to a constraint on the utility/usefulness of the released data.

In statistical data mining, unknown quantities are conventionally treated as random variables, measurable functions that map the probability space to the event space. As a result, privacy is usually measured as a stochastic feature using, e.g., the mutual information in [7] or the Rényi divergence[1] in [4], [9]. However, in some practical applications, it may be preferred to measure and optimize privacy or information leakage in a non-stochastic setting. For instance, consider the problem of releasing a table of just 50 records. The size is not large enough for making probabilistic inference about the population[2]. The data curator's concern could be whether an adversary can gain private information that is not based on statistics. Kolmogorov had partly answered this question in his early study [11] stating that the more values of sensitive data appears concurrently with the released data, the less certain an adversary is about the sensitive data. This is also an information-theoretic interpretation of the k-anonymity [12] and motivated the adoption of non-stochastic privacy metrics [13], [14]. Non-stochastic treatment of variables and measures of information have proved popular in other engineering domains, such as networked control and estimation in which variables might be bounded but do not follow a well-defined probability distribution [15]–[18].

In this paper, we study how to preserve privacy against non-stochastic inference, where the adversary is able to deduce non-statistical features of the data while both privacy and utility are measured by non-probabilistic information quantities. Specifically, for a data curator who wants to share non-

The authors are with the University of Melbourne.
emails:{ni.ding,farhad.farokhi}@unimelb.edu.au

The work of Ni Ding is funded by the Doreen Thomas Postdoctoral Fellowship at the University of Melbourne.

The work of F. Farokhi is funded by the Melbourne School of Engineering at the University of Melbourne.

[1]DP can be defined as the supremum of the Rényi divergence $D_\alpha(p(\cdot)\|q(\cdot))$ for $\alpha = \infty$ [10].

[2]This could be the case in a small experiment or at the beginning of data release in a gradual data release setting. For example, in small datasets, each row of a tabular dataset is often distinct and the information only depends on the range, the number of distinct values, and not probabilities or statistics. We show an example of this situation in Section VII-B.

sensitive data X with the public, but to protect the sensitive data S that is related to X, the problem can be cast as generating a sanitization $\hat{X}$ with a specified level of the data fidelity/utility while leaking the least amount of information about S. Based on non-stochastic information theory, two privacy measures are considered: $L_0(S \to \hat{X})$ capturing the information on S conveyed by $\hat{X}$ [13] and the *maximin information* $I_*(S;\hat{X})$ measuring the adversary's knowledge on S refined by $\hat{X}$ [15]. We reveal different interpretations in privacy between these two measures: $L_0(S \to \hat{X})$ measures the maximum uncertainty reduction on S at the adversary side, which is shown to correspond to k-anonymity, while $I_*(S;\hat{X})$ measures how distinguishable S can be by observing $\hat{X}$, or how much private information can be obtained without error by the adversary.

We consider the privacy-preserving problem by generating $\hat{X}$ that minimizes the privacy leakage, $L_0(S \to \hat{X})$ or $I_*(S;\hat{X})$ subject to maintaining the utility $U(X;\hat{X})$ of the released data above some threshold. The non-stochastic sanitization $\hat{X}$ is done by a quantization method, or a deterministic clustering of the alphabet of the public data X. To this end, we propose a greedy clustering algorithm for extracting a locally optimal solution $\hat{X}$ by iteratively merging two elements of X that strictly reduces the Lagrangian function of the privacy-preserving problem. We show that the value of the maximin information $I_*(S;X)$ is determined by the maximum number of disconnected subgraphs in an undirected uncapacitated graph, which is equivalent to the *confusability graph* for the study on Shannon's zero-error capacity [19]. Therefore, the proposed agglomerative clustering algorithm for minimizing $I_*(S;\hat{X})$ is analogous to a subgraph merging process. To instantiate the solution to the problem, we study two specific utility measures for $U(X;\hat{X})$: the minimum uncertainty reduction $I_0(X \to \hat{X})$, the worst-case resolution of the public data X via the released data $\hat{X}$ [15], and the maximum distortion $\max_{x,\hat{x}} d(x,\hat{x})$ of the sanitized data $\hat{X}$. We also investigate the relationship between our work and the stochastic information-theoretic privacy studies in [7]–[9]. We elaborate that the Gács-Körner common information [20] is the stochastic version of the maximin information, indicating when statistical indistinguishability, e.g., the DP [4], is attainable.

Finally, we show how to apply the proposed agglomerative clustering algorithm to plot the Pareto frontier, where one can directly search for a privacy-preserving data sanitization solution for any given utility constraint. We also run experiments on a real-world dataset to show that the proposed agglomerative clustering algorithm is capable of realistic/practical privacy-preserving sanitization.

Related works: For a continuous sensitive variable S, the problem of designing the quantization $f(S)$ and its sanitization $\hat{f}(S)$ to minimize the privacy leakage $L_0(S \to \hat{f}(S))$ or $I_*(S;\hat{f}(S))$ subject to an upper bound on the ℓ_2-norm $\|f(S)-\hat{f}(S)\|_2$ was investigated in [13]. This reduces to the privacy preserving problem in this paper when $X = f(S)$. This paper generalizes the framework of [13] as there is no requirement that X is a deterministic function of S. Moreover, the difference between $L_0(S \to \hat{f}(S))$ and $I_*(S;\hat{f}(S))$, the graph decomposition method for determining $I_*(S;f(S))$ and its role in minimizing $I_*(S;\hat{f}(S))$ are not discussed in [13]. The setup of this paper matches the typical setting in statistical inference[3] and is adopted in most of the stochastic information-theoretic privacy studies [7]–[9] (see Section VI).

Organization: This paper is organized as follows. Section II reviews some definitions in non-stochastic information theory. Section III introduces privacy measures L_0 and I_*. Section IV proposes a graph decomposition method for determining I_*. In Section V, we formulate the non-stochastic privacy-preserving problem and propose two agglomerative clustering algorithms. Section VI studies the relationship between our work and the stochastic information-theoretic privacy. Section VII presents the experimental results.

II. Non-stochastic Information Theory

In non-stochastic information theory [15], for the sample space Ω, we say that the mapping $X : \Omega \to \mathbb{X}$ is an *uncertain variable (uv)*. For a sample $\omega \in \Omega$, $X(\omega)$ denote a realization of the uv X. The notation $X(\omega)$ is simplified to X in the rest of the paper if the context is clear. When $\mathbb{X}$ is a finite set, X is a *discrete uv*. In this paper, we only consider discrete uvs. Here, a uv differs from the a random variable (rv) in probability theory in that we do not assume a σ-algebra subset family over the sample space Ω or a probability measure over this subset family. For a pair of discrete uvs S and X, let $[\![S,X]\!] \triangleq \{(S(\omega),X(\omega))\colon \omega\in\Omega\}$ and $[\![S|x]\!] \triangleq \{S(\omega)\colon X(\omega)=x, \omega\in\Omega\}$, respectively, be the joint range of S and X and the conditional range of S based on the observation that $X=x$. Note, we can rewrite $[\![S,X]\!] = \bigcup_{x\in[\![X]\!]} \big([\![S|x]\!]\times\{x\}\big)$ and $[\![S|X]\!] = \{[\![S|x]\!]\colon x\in[\![X]\!]\}$. We say that S and X are *independent*, if $[\![S|x]\!] = [\![S]\!], \forall x\in[\![X]\!]$. For $[\![S,X]\!]$, we have the marginal ranges $[\![S]\!]$ and $[\![X]\!]$. The non-probabilistic entropy of a uv X is defined as

$$H_0(X) \triangleq \log|[\![X]\!]|,$$

which coincides with the Hartley (maximum) entropy [21] or the Rényi entropy H_α in the case $\alpha=0$ [22]. We assume the base of all logarithms is 2 in this paper.

III. Non-stochastic Inference and Leakage

The sensitive/private data is denoted by S. Consider the case where a data curator wants to release X to the public, which is related[4] with S. For example, the taxation office may want to release the income records, which, even if anonymized, could be observed by an adversary in the public domain to infer individuals' identities. The data curator needs to maintain a certain level of usefulness/utility of the released data, e.g., to the ensure correctness of answers to legitimate surveys/queries. At the same time, the curator has privacy concerns in the sense that the more fidelity the released data has, the easier

[3]Statistical inference considers the situation where a user/adversary is planning to deduce properties S of a population X based on noisy observations $\hat{X}$, where the statistics of the population X is usually given or specified.

[4]Relatedness, similar to correlation in the probabilistic sense, states that $[\![S|x]\!] = [\![S]\!]$ does not hold for all $x\in[\![X]\!]$, or S and X are not non-stochastically independent uvs.

an adversary can infer the private data S. Thus, it is necessary for the curator to be equipped with a valid privacy measure to anticipate the risk of the information leakage.

A. Uncertainty Reduction

For any realization x of the public data X, Kolmogorov defined a 'combinatorial' conditional entropy as $\log [\![S|x]\!]$ in [11]. This interprets the uncertainty/entropy reduction $\log(|[\![S]\!]|/|[\![S|x]\!]|)$ as a measure of non-stochastic information on S gained by an adversary in the public domain after observing x, or the privacy leakage at the data curator side by releasing x. For the conditional information in the sense of the maximum entropy $H_0(S|X) \triangleq \max_{x\in[\![X]\!]} \log |[\![S|x]\!]|$, the author in [15] proposed the *non-stochastic* 0*-information* as the minimal difference between prior and posterior entropy

$$I_0(S \to X) \triangleq H_0(S) - H_0(S|X) = \min_{x\in[\![X]\!]} \log \frac{|[\![S]\!]|}{|[\![S|x]\!]|}.$$

However, in [13], it was suggested to consider the worst-case, i.e., the minimal posterior entropy/uncertainty $B_0(S|X) \triangleq \min_{x\in[\![X]\!]} \log |[\![S|x]\!]|$, to quantify the *maximal information leakage*

$$L_0(S \to X) \triangleq H_0(S) - B_0(S|X) = \max_{x\in[\![X]\!]} \log \frac{|[\![S]\!]|}{|[\![S|x]\!]|}.$$

as a measure of privacy.

B. Maximin Information/Indistinguishability

Note that neither I_0 or L_0 is symmetric. That is, $I_0(S \to X) = I_0(X \to S)$ or $L_0(S \to X) = L_0(X \to S)$ do not hold in general. However, a symmetric measure of the dependence between S and X can be defined based on the concept of overlap partition.

Definition 1 (Overlap Partition [15, Definitions 3.1 and 3.2, Lemma 3.1])**.**

(a) ***overlap connectedness****: A pair $s, s' \in [\![S]\!]$ is $[\![S|X]\!]$-overlap connected, denoted by $s \leftrightsquigarrow s'$, if there exists an ordered finite sequence $(x_1, \ldots, x_n) \in [\![X]\!]^n$ such that $s \in [\![S|x_1]\!]$, $s' \in [\![S|x_n]\!]$ and $[\![S|x_i]\!] \cap [\![S|x_{i+1}]\!] \neq \emptyset$ for all $i \in \{1, \ldots, n-1\}$. Thus, $s \leftrightsquigarrow s, \forall s \in [\![S]\!]$ and $s \leftrightsquigarrow s', \forall s, s' \in [\![S|x]\!]$, $x \in [\![X]\!]$. A subset $\mathcal{S} \subseteq [\![S]\!]$ is $[\![S|X]\!]$-overlap connected if $s \leftrightsquigarrow s', \forall s, s' \in \mathcal{S}$;*

(b) ***overlap isolation****: Two subsets $\mathcal{S}, \mathcal{S}' \subseteq [\![S]\!]$ are $[\![S|X]\!]$-overlap isolated if there does not exist $s \in \mathcal{S}$ and $s' \in \mathcal{S}'$ such that $s \leftrightsquigarrow s'$;*

(c) ***overlap isolated partition****: A partition of $[\![S]\!]$ is a $[\![S|X]\!]$-overlap isolated partition, denoted by $\mathcal{P}_{[\![S|X]\!]}$, if every distinct $\mathcal{S}, \mathcal{S}' \in \mathcal{P}_{[\![S|X]\!]}$ are $[\![S|X]\!]$-overlap isolated.*

(d) ***overlap partition****: A $[\![S|X]\!]$-overlap isolated partition $\mathcal{P}_{[\![S|X]\!]}$ is called $[\![S|X]\!]$-overlap partition if each $\mathcal{S} \in \mathcal{P}_{[\![S|X]\!]}$ is $[\![S|X]\!]$-overlap connected;*

(e) ***maximin information****: Let $\mathcal{P}^*_{[\![S|X]\!]}$ be the finest $[\![S|X]\!]$-overlap partition with the largest size $|\mathcal{P}^*_{[\![S|X]\!]}|$. Such partition is unique. The maximin information is defined as*

$$I_*(S;X) = \log |\mathcal{P}^*_{[\![S|X]\!]}|.$$

The maximin information $I_*(S;X)$ measures the 'refined' knowledge [15, Section III-B], or the highest resolution, on the range $[\![S]\!]$ by observing X. We show in the next subsection that $I_*(S;X)$ indicates the non-stochastic distinguishability in the view of privacy.

C. $L_0(S \to X)$, $I_*(S;X)$, k-anonymity and Non-stochastic Distinguishability

For the release of tabular data, k-*anonymity* was proposed in [12] to guarantee that an adversary cannot distinguish between at least k rows of the record for each instance of the released data.

Definition 2 (k-anonymity)**.** *$[\![S|X]\!]$ is k-anonymous if $|[\![S|x]\!]| \geq k$ for all $x \in [\![X]\!]$.*

The method for attaining k-anonymity is to ensure k-occurrences of the records $|[\![S|x]\!]| = k$ for each x in the released table.[5] A straightforward result from Definition 2 is the one-to-one correspondence between L_0 and the k-anonymity.

Lemma 1 ($L_0 \iff k$-anonymity)**.** *$[\![S|X]\!]$ is k-anonymous if and only if $B_0(S|X) \geq \log k$ or $L_0(S \to X) \leq \log(|[\![S]\!]|/k)$.* ■

While L_0 or the k-anonymity denotes the non-stochastic uncertainty reduction, $I_*(S;X)$ measures the non-stochastic distinguishability. In [14], the non-stochastic privacy is measured as the change of the range of the public data $|[\![X|s]\!] \setminus [\![X|s']\!] \sqcup [\![X|s']\!] \setminus [\![X|s]\!]|$ conditioned on two distinct sensitive uv values s and s'. Here, $|[\![X|s]\!] \setminus [\![X|s']\!] \sqcup [\![X|s']\!] \setminus [\![X|s]\!]|$ measures the distinguishability between s and s' (the lower the value, the less distinguishable and the more privacy). This can be viewed as a non-stochastic version of the differential privacy (DP) [4]. Clearly, if $I_*(S;X) > 0$, there are at least two non-overlapping/disjoint $\mathcal{S}, \mathcal{S}' \subsetneq \mathcal{P}^*_{[\![S|X]\!]}$ such that an adversary can discriminate perfectly any pair of $s \in \mathcal{S}$ and $s' \in \mathcal{S}'$ by observing X.[6] In this case, any distinguishability-based privacy measure is minimized (no privacy). It also relates to the zero-error capacity [24]: $I_*(S;X)$ is the largest information amount (in bits) of S that can be conveyed without any error by transmitting X [15, Theorem 4.1]. In addition, we can prove the following relationship between I_* and L_0. The proof is presented in Appendix A.

Proposition 1. $I_*(S;X) \leq L_0(S \to X)$. ■

This means that $L_0(S \to X)$ could be nonzero when $0 = I_*(S;X)$, i.e., even if the adversary cannot perfectly obtain some bits of S from X, he/she might still become more certain about S. Apart from these non-stochastic measures, we show

[5]The value k in the original definition [12, Definition 3] refers to the number of records, including k records with the same value of the sensitive data s such that $s \in [\![S|x]\!]$. It is shown in [13], [23] that a k-anonymous database with the identical occurrences of the sensitive variable is subject to the homogeneity attack [13], [23]. This does not apply to the k-anonymity in Definition 2, which is a more strict definition where k refers to the number of distinct values of S.

[6]This is because $|[\![X|s]\!] \setminus [\![X|s']\!] \sqcup [\![X|s']\!] \setminus [\![X|s]\!]|$ attains the maximum if $[\![X|s]\!] \setminus [\![X|s']\!] = [\![X|s]\!]$ and $[\![X|s']\!] \setminus [\![X|s]\!] = [\![X|s']\!]$.

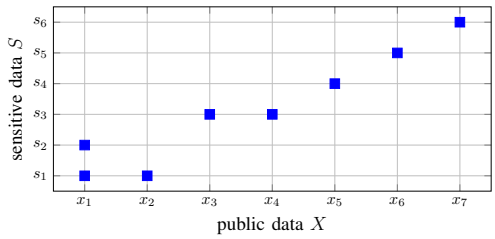


Fig. 1. The joint range $[\![S,X]\!]$ of uvs S and X with $[\![S]\!]=\{s_1,\ldots,s_6\}$ and $[\![X]\!]=\{x_1,\ldots,x_7\}$. A blue square at (s_i,x_j) denotes $(s_i,x_j)\in[\![S,X]\!]$; otherwise $(s_i,x_j)\notin[\![S,X]\!]$.

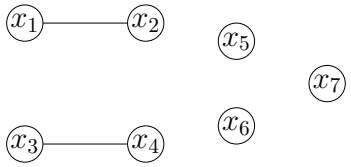


Fig. 2. The undirected graph $G_X=([\![X]\!],\mathcal{E}_{[\![S|X]\!]})$ based on the conditional range $[\![S|X]\!]$ of Fig. 1. Both the min-cut and graph connectivity of G_X is 0, i.e., G_X is disconnected, which means $I_*(S;X)>0$. The finest decomposition of G_X is $\mathcal{P}^*_{G_X}=\{\{x_1,x_2\},\{x_3,x_4\},\{x_5\},\{x_6\},\{x_7\}\}$, which results in $I_*(S;X)=\log|\mathcal{P}^*_{G_X}|=\log 5$.

in Section VI how L_0 and I_* relate to stochastic information theory.

IV. Graph Decomposition for Determining $I_*(S;X)$

Let $G_X=([\![X]\!],\mathcal{E}_{[\![S|X]\!]})$ be an undirected uncapacitated graph with the node set $[\![X]\!]$ and the edge set $\mathcal{E}_{[\![S|X]\!]}=\{(x,x')\colon [\![S|X=x]\!]\cap[\![S|X=x']\!]\neq\emptyset\}$. Here, G_X corresponds to the *adjacency matrix* in stochastic information theory [24] $A=[a_{x,x'}]$ where $a_{x,x'}=1$ if $\exists s\in[\![S]\!]$ such that the joint probabilities $p(s,x)$ and $p(s,x')$ are both nonzero and $a_{x,x'}=0$ otherwise. Therefore, G_X is also called the *confusability graph* [19].[7]

A *decomposition* $\mathcal{P}_{G_X}$ of G_X is a partition of $[\![X]\!]$ such that any two distinct subgraphs $\mathcal{X},\mathcal{X}'\in\mathcal{P}_{G_X}$ are disconnected; The *finest decomposition*, denoted by $\mathcal{P}^*_{G_X}$, is a decomposition such that each subgraph $\mathcal{X}$ is connected [27]. The following lemma states the equivalence between graph decomposition and the $[\![S|X]\!]$-overlap partition. The proof is in Appendix B.

Lemma 2. *For each decomposition $\mathcal{P}_{G_X}$, $\{[\![S|\mathcal{X}]\!]:\mathcal{X}\in\mathcal{P}_{G_X}\}$ is a $[\![S|X]\!]$-overlap isolated partition; For the finest decomposition $\mathcal{P}^*_{G_X}$, $\mathcal{P}^*_{[\![S|X]\!]}=\{[\![S|\mathcal{X}]\!]:\mathcal{X}\in\mathcal{P}^*_{G_X}\}$ is the unique $[\![S|X]\!]$-overlap partition so that the maximin information is $I_*[S;X]=\log|\mathcal{P}^*_{G_X}|$.* ■

For $\mathcal{X}\subseteq[\![X]\!]$, let $\kappa(\mathcal{X})=|\{(x,x')\in\mathcal{E}_{[\![S|X]\!]}:x\in\mathcal{X},x'\in[\![X]\!]\setminus\mathcal{X}\}|$ denote the value of the cut $\{\mathcal{X},[\![X]\!]\setminus\mathcal{X}\}$. The graph G_X is decomposable if the min-cut $\min\{\kappa(\mathcal{X}):\emptyset\neq\mathcal{X}\subsetneq[\![X]\!]\}=0$. The solution to the min-cut problem $\arg\min\{\kappa(\mathcal{X}):\emptyset\neq\mathcal{X}\subsetneq[\![X]\!]\}$ forms a set lattice [28, Section 2.3][8] and can be determined by the max-flow algorithm [29] in polynomial time. All the smallest min-cut solutions constitute the finest decomposition [28, Section 2.2]:

$$\mathcal{P}^*_{G_X}=\bigcap\arg\min\{\kappa(\mathcal{X}):\emptyset\neq\mathcal{X}\subsetneq[\![X]\!]\}.$$

A simple method to determine $\mathcal{P}^*_{G_X}$ is to recursively run either breadth-first search (BFS) or death-first search (DFS) [30]. We run BFS or DFS staring from any node $x\in X$ to find all other nodes that are reachable (via some path) from x. This identifies a connected subgraph in $\mathcal{P}^*_{G_X}$. Then, we repeat this procedure for any node that has not been searched until no such node is left. The complexity of BFS/DFS is $O(|[\![X]\!]|+|\mathcal{E}_{[\![S|X]\!]}|)$. The above method calls BFS/DFS for $|\mathcal{P}^*_{G_X}|$ times and therefore the complexity is upper bounded by $O(|[\![X]\!]|^2+|[\![X]\!]||\mathcal{E}_{[\![S|X]\!]}|)$.

There are also other methods that determine $\mathcal{P}^*_{G_X}$. For example, one of algorithms in [27], [31], [32] for determining the network strength.[9] The lowest complexity of these algorithms is $O(|[\![X]\!]|)$ runs of the max-flow algorithm. Without explicitly constructing the graph G_X, we can still obtain the Laplacian of the confusability matrix to determine the finest decomposition $\mathcal{P}^*_{G_X}$; see [33, Remark 6]. Note that the validity of this Laplacian method is due to the inherit relationship between the maximin information I_* and Gács-Körner common information, which is explained in Section VI.

Example 1. *For $[\![S]\!]=\{s_1,\ldots,s_6\}$, $[\![X]\!]=\{x_1,\ldots,x_7\}$ and the joint range $[\![S,X]\!]$ shown in Fig. 1, we have $G_X=([\![X]\!],\mathcal{E}_{[\![S|X]\!]})$ in Fig. 2 with the min-cut being 0, i.e., the graph is decomposable. The finest decomposition $\mathcal{P}^*_{G_X}=\{\{x_1,x_2\},\{x_3,x_4\},\{x_5\},\{x_6\},\{x_7\}\}$ and the set of all min-cut solutions can be constructed by the fusion of subsets in $\mathcal{P}^*_{G_X}$:*

$$\arg\min\{\kappa(\mathcal{X}):\emptyset\neq\mathcal{X}\subsetneq[\![X]\!]\}=\left\{\bigsqcup_{\mathcal{X}\in\mathfrak{X}}\mathcal{X}\colon\mathfrak{X}\subseteq\mathcal{P}^*_{G_X}\right\}\setminus\{[\![X]\!]\}.$$

For example, $\{x_1,\ldots,x_4\}=\{x_1,x_2\}\sqcup\{x_3,x_4\}$ is a min-cut solution and so is $\{x_5,x_6,x_7\}=\{x_5\}\sqcup\{x_6\}\sqcup\{x_7\}$. The finest decomposition $\mathcal{P}^*_{G_X}$ determines the unique $[\![S|X]\!]$-overlap partition $\mathcal{P}^*_{[\![S|X]\!]}=\{[\![S|\mathcal{X}]\!]\colon\mathcal{X}\in\mathcal{P}^*_{G_X}\}=\{\{s_1,s_2\},\{s_3\},\{s_4\},\{s_4\},\{s_6\}\}$ and thus $I_*(S;X)=\log|\mathcal{P}^*_{[\![S|X]\!]}|=\log|\mathcal{P}^*_{G_X}|=\log 5$.

[7]The adjacency matrix A in [24] and the corresponding confusability graph in [19] only depend on whether the joint probability $p(s,x)$ is nonzero or not. By knowing that each non-zero $p(s,x)$ determines the presence of (s,x) in $[\![S,X]\!]$, the stochastic joint probability $p(s,x),\forall s,x$ can be reduced to the non-stochastic joint range by $[\![S,X]\!]=\{(s,x)\colon p(s,x)\neq 0\}$. This explains that both the Shannon capacity of a graph, the zero-error capacity, and the maximin information $I_*(S,X)$ are related to the graph theory or combinatorics [25], [26].

[8]More precisely, the set $\mathcal{T}=\arg\min\{\kappa(\mathcal{X}):\emptyset\neq\mathcal{X}\subsetneq[\![X]\!]\}\sqcup\{\emptyset\}\sqcup\{[\![X]\!]\}$ is a lattice such that, for any $\mathcal{X}_1,\mathcal{X}_2\in\mathcal{T}$, $\mathcal{X}_1\cap\mathcal{X}_2\in\mathcal{T}$ and $\mathcal{X}_1\cup\mathcal{X}_2\in\mathcal{T}$.

[9]If the network strength is nonzero, G_X is connected and $I(S;X)=0$. However, if the network strength is zero, G_X is decomposable and the algorithms in [27], [31], [32] return the finest decomposition $\mathcal{P}^*_{G_X}$ [27].

V. Non-stochastic Privacy-Preserving Data Release

Instead of the original X, the data curator publishes $\hat{X}$, a distorted or sanitized version of the public data X. The problem is how to sanitize the data to preserve privacy, while keeping the utility of $\hat{X}$ above a certain level. In this section, we formulate the non-stochastic privacy-preserving problem as the minimization of either of two privacy measures in Section III, $L_0(S \to \hat{X})$ or $I_*(S;\hat{X})$, subject to a utility constraint. We consider the Lagrangian function of the non-stochastic privacy-preserving problem and propose two agglomerative clustering algorithms to determine $\hat{X}$ for the minimization of $L_0(S \to \hat{X})$ and $I_*(S;\hat{X})$, respectively. We also consider the problem of finding a privacy-preserving data release scheme $\hat{X}$ that guarantees the non-stochastic indistinguishability, which is formulated as minimizing $L_0(S \to \hat{X})$ subject to $I_*(S;\hat{X}) = 0$ and the utility constraint, and show that this problem can also be solved by the proposed agglomerative clustering algorithm.

The main objective of the non-stochastic privacy-preserving problem is to determine $[\![\hat{X}]\!]$ and the conditional range $[\![\hat{X}|X]\!]$ through a deterministic function $f\colon [\![X]\!] \mapsto [\![\hat{X}]\!]$ with $|[\![X]\!]| \geq |[\![\hat{X}]\!]|$. Here, f can be considered as a quantization function that clusters/merges all $x \in [\![X|\hat{x}]\!]$ to $\hat{x} = f(x)$. We use a more convenient notation for f, the partition $\mathcal{Q} = \{[\![X|\hat{x}]\!]\colon \hat{x} \in [\![\hat{X}]\!]\}$. This privacy-preserving method is captured by the Markov chain $S - X - \hat{X}$, where, for a given joint range $[\![S,X]\!]$, a sanitization method $[\![\hat{X}|X]\!]$ results in the conditional range

$$[\![S|\hat{x}]\!] = \bigcup_{x \in [\![X|\hat{x}]\!]} [\![S|x]\!] = [\![S|\mathcal{X}]\!], \quad \forall \hat{x} \in [\![\hat{X}]\!] \tag{1}$$

for $\mathcal{X} = [\![X|\hat{x}]\!] \in \mathcal{Q}$.

A. Minimizing Privacy Leakage $L_0(S \to \hat{X})$

For a given threshold θ, consider the problem

$$\min_{[\![\hat{X}|X]\!]} L_0(S \to \hat{X}), \quad \text{s.t. } U_i(X;\hat{X}) \geq \theta. \tag{2}$$

That is, we aim to maximize the value of k in k-anonymity (Lemma 1) subject to a constraint on the utility. For the quantization $\mathcal{Q}$ corresponding to $[\![\hat{X}|X]\!]$, we have $B_0(S|\hat{X}) = \min_{\hat{x} \in [\![\hat{X}]\!]} \log |[\![S|\hat{x}]\!]| = \min_{\mathcal{X} \in \mathcal{Q}} \log |[\![S|\mathcal{X}]\!]|$. The Lagrangian function of (2) is $\mathcal{L}([\![\hat{X}|X]\!], \lambda) = -\min_{\mathcal{X} \in \mathcal{Q}} \log |[\![S|\mathcal{X}]\!]| - \lambda U_i(X;\hat{X})$ for all $\lambda \in [0, +\infty)$.

The utility of the released data $\hat{X}$ is measured by $U_i(X;\hat{X})$. We consider two definitions:

$$U_i(X;\hat{X}) = \begin{cases} I_0(X \to \hat{X}), & i = 1, \\ -\max_{x,\hat{x}\colon \hat{x} \in [\![\hat{X}|x]\!]} d(x,\hat{x}), & i = 2. \end{cases}$$

The non-stochastic 0-information $I_0(X \to \hat{X}) = H_0(X) - \max_{\hat{x} \in [\![\hat{X}]\!]} \log |[\![X|\hat{x}]\!]|$ denotes the minimal posterior entropy reduction on X, which corresponds to coarsest resolution of X by observing $\hat{X}$. Therefore, $I_0(X \to \hat{X})$ indicates the worst-case data utility. So is $-\max_{x,\hat{x}\colon \hat{x} \in [\![\hat{X}|x]\!]} d(x,\hat{x})$ for the pairwise distance/distortion[10] function $d(\cdot,\cdot)$.

Algorithm 1: Agglomerative clustering algorithm for solving problem (2)

input : the Lagrangian multiplier $\lambda \in [0, +\infty)$ and the joint range $[\![S,X]\!]$.
output: the quantization $\mathcal{Q}^{(t-1)}$.

1 initiate $\mathcal{Q}^{(0)} \coloneqq \{\{x\}\colon x \in [\![X]\!]\}$ and $t \coloneqq 0$;
2 **repeat**
3 $\quad \Pi \coloneqq \arg\min\{\log |[\![S|\mathcal{X}]\!]|\colon \mathcal{X} \in \mathcal{Q}^{(t)}\}$;
4 $\quad \mathcal{Q}^{(t+1)} \coloneqq \mathcal{Q}^{(t)}$;
5 $\quad t \coloneqq t+1$;
6 $\quad$ **repeat**
7 $\quad\quad \mathcal{X}^* \coloneqq \arg\max \{U_i(X;\hat{X}^{(t)}_{\{\mathcal{X},\mathcal{X}'\}})\colon \mathcal{X}' \in \mathcal{Q}^{(t)}, \mathcal{X}' \neq \mathcal{X}, [\![S|\mathcal{X}']\!] \neq [\![S|\mathcal{X}]\!]\}$;
9 $\quad\quad \mathcal{Q}^{(t)} \coloneqq \mathcal{Q}^{(t)}_{\{\mathcal{X},\mathcal{X}^*\}}$;
10 $\quad\quad \Pi \coloneqq \Pi \setminus \{\mathcal{X}, \mathcal{X}^*\}$;
11 $\quad$ **until** $\Pi = \emptyset$;
12 $\quad \triangle\mathcal{L} \coloneqq L([\![\hat{X}^{(t)}|X]\!], \lambda) - L([\![\hat{X}^{(t-1)}|X]\!], \lambda)$;
13 **until** $\triangle\mathcal{L} \geq 0$ *or* $|\mathcal{Q}^{(t)}| = 1$;
14 **return** $\mathcal{Q}^{(t-1)}$;

1) Agglomerative Clustering Algorithm: While problem (2) is a computationally complex optimization problem, in particular for large datasets, we propose a greedy agglomerative clustering in Algorithm 1 that generates a locally optimal $[\![\hat{X}]\!]$ and $[\![\hat{X}|X]\!]$ by iteratively merging the elements in $[\![X]\!]$.

Algorithm 1 starts with $\mathcal{Q}^{(0)} = \{\{x\}\colon x \in [\![X]\!]\}$, which means no quantization. In each iteration t, we strictly increase $\min_{\mathcal{X} \in \mathcal{Q}^{(t)}} \log |[\![S|\mathcal{X}]\!]|$ by merging each $\mathcal{X} \in \arg\min\{\log |[\![S|\mathcal{X}]\!]|\colon \mathcal{X} \in \mathcal{Q}^{(t)}\}$ with another subset $\mathcal{X}^*$ that maximize the utility function $U_i(X;\hat{X}^{(t)}_{\{\mathcal{X},\mathcal{X}'\}})$ over the subsets $\mathcal{X}'$ in the current quantization $\mathcal{Q}^{(t)}$. We explain how to search $\mathcal{X}^*$ in step 7 as follows.

For each distinct pair $\mathcal{X}_1, \mathcal{X}_2 \in \mathcal{Q}^{(t)}$, let the partition

$$\mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}} = (\mathcal{Q}^{(t)} \setminus \{\mathcal{X}_1, \mathcal{X}_2\}) \cup \{\mathcal{X}_1 \sqcup \mathcal{X}_2\}$$

be obtained by merging $\mathcal{X}_1$ and $\mathcal{X}_2$ and $\hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}$ is the resulting uv. Consider the utility function $U_1(X;\hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}) = I_0(X \to \hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}) = H_0(X) - \max_{\mathcal{X} \in \mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}} \log |\mathcal{X}|$. Because merging $\mathcal{X}_1$ and $\mathcal{X}_2$ such that $[\![S|\mathcal{X}_1]\!] = [\![S|\mathcal{X}_2]\!]$ does not change the value of $L_0(S \to \hat{X}^{(t)})$, we let

$$\Phi(\mathcal{Q}^{(t)}) = \{(\mathcal{X}_1, \mathcal{X}_2) \in \mathcal{Q}^{(t)}\colon \mathcal{X}_1 \neq \mathcal{X}_2, [\![S|\mathcal{X}_1]\!] \neq [\![S|\mathcal{X}_2]\!]\}$$

and consider

$$\begin{aligned}(\mathcal{X}_1^*, \mathcal{X}_2^*) &\in \operatorname*{arg\,max}_{(\mathcal{X}_1,\mathcal{X}_2) \in \Phi(\mathcal{Q}^{(t)})} U_1(X;\hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}) \\ &= \operatorname*{arg\,min}_{(\mathcal{X}_1,\mathcal{X}_2) \in \Phi(\mathcal{Q}^{(t)})} \max_{\mathcal{X} \in \mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}} \log |\mathcal{X}|\end{aligned}$$

[10] $d(\cdot,\cdot)$ could be the ℓ_p-norm or any other pairwise dissimilarity measure.

Here, we have $\max_{\mathcal{X}\in\mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}} \log|\mathcal{X}| = \log\big(\max\{|\mathcal{X}_1|+|\mathcal{X}_2|, \max_{\mathcal{X}\in\mathcal{Q}^{(t)}}|\mathcal{X}|\}\big)$, for which, if $|\mathcal{X}_1|+|\mathcal{X}_2| \le \max_{\mathcal{X}\in\mathcal{Q}^{(t)}}|\mathcal{X}|, \forall(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})$, then

$$\operatorname*{arg\,min}_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})} \max_{\mathcal{X}\in\mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}} \log|\mathcal{X}| = \Phi(\mathcal{Q}^{(t)});$$

if there exists $(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})$ such that $|\mathcal{X}_1|+|\mathcal{X}_2| > \max_{\mathcal{X}\in\mathcal{Q}^{(t)}}|\mathcal{X}|$, then

$$\operatorname*{arg\,min}_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})} \max_{\mathcal{X}\in\mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}} \log|\mathcal{X}| = \operatorname*{arg\,min}_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})} \{|\mathcal{X}_1|+|\mathcal{X}_2|\}.$$

Therefore, we have $\arg\min_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})}\{|\mathcal{X}_1|+|\mathcal{X}_2|\} \subseteq \arg\min_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})} \max_{\mathcal{X}\in\mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}} \log|\mathcal{X}|$. So, we can search for

$$(\mathcal{X}_1^*,\mathcal{X}_2^*) \in \operatorname*{arg\,min}_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})} \big\{|\mathcal{X}_1|+|\mathcal{X}_2|\big\}.$$

Therefore, for the utility function $U_1(X;\hat{X}^{(t)}_{\{\mathcal{X},\mathcal{X}'\}}) = I_0(X\to\hat{X}^{(t)}_{\{\mathcal{X},\mathcal{X}'\}})$, we choose $\mathcal{X}^* \in \arg\min\big\{|\mathcal{X}\sqcup\mathcal{X}'|\colon \mathcal{X}'\in\mathcal{Q}^{(t)}, \mathcal{X}'\neq\mathcal{X}, [\![S|\mathcal{X}']\!]\neq[\![S|\mathcal{X}]\!]\big\}$.

For the utility function $U_2 = -\max_{x,\hat{x}\colon \hat{x}\in[\![\hat{X}|x]\!]} d(x,\hat{x})$, a unique codeword $\hat{x}$ is assigned to each cluster $\mathcal{X}\in\mathcal{Q}$ such that $\mathcal{X} = [\![X|\hat{x}]\!]$. Two possible choices are the centroid of the cluster $\hat{x} = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} x$ and $\hat{x} = x$ for some $x\in\mathcal{X}$, but other choices are possible based on the application. We denote the maximum distortion for the cluster $\mathcal{X}\in\mathcal{Q}$ by

$$\bar{d}(\mathcal{X}) = \max_{x\in\mathcal{X}} d(x,\hat{x})$$

and have

$$(\mathcal{X}_1^*,\mathcal{X}_2^*) \in \operatorname*{arg\,min}_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})} \max_{x,\hat{x}\colon \hat{x}\in[\![\hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}|x]\!]} d(x,\hat{x}) = \operatorname*{arg\,min}_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})} \max_{\mathcal{X}\in\mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}} \bar{d}(\mathcal{X})$$

Here, since $\max_{\mathcal{X}\in\mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}} \bar{d}(\mathcal{X}) = \max\{\bar{d}(\mathcal{X}_1\sqcup\mathcal{X}_2), \max_{\mathcal{X}\in\mathcal{Q}^{(t)}}\bar{d}(\mathcal{X})\}$, $\arg\min_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})}\bar{d}(\mathcal{X}_1\sqcup\mathcal{X}_2) \subseteq \arg\min_{(\mathcal{X}_1,\mathcal{X}_2)\in\Phi(\mathcal{Q}^{(t)})}\max_{\mathcal{X}\in\mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}}\bar{d}(\mathcal{X})$ and therefore we choose $\mathcal{X}^*\in\arg\min\big\{\bar{d}(\mathcal{X}\sqcup\mathcal{X}')\colon \mathcal{X}'\in\mathcal{Q}^{(t)}, \mathcal{X}'\neq\mathcal{X}, [\![S|\mathcal{X}']\!]\neq[\![S|\mathcal{X}]\!]\big\}$ in step 7.

Example 2. *For $\lambda = 0.3$, we run Algorithm 1 to the uvs S and X in Example 1 with the initiation $\mathcal{Q}^{(0)} = \{\{x_1\},\dots,\{x_7\}\}$. Consider the utility function $U_1(X;\hat{X}^{(t)}) = I_0(S\to\hat{X}^{(t)})$. At the first iteration, we have $\Pi = \{\{x_2\},\dots,\{x_7\}\}$, for which step 2 to step 11 produce the quantization $\mathcal{Q}^{(1)} = \{\{x_1,x_2,x_7\},\{x_3,x_5\},\{x_4,x_6\}\}$ with the conditional ranges $[\![S|\{x_1,x_2,x_7\}]\!] = \{s_1,s_2,s_6\}$, $[\![S|\{x_3,x_5\}]\!] = \{s_3,s_4\}$ and $[\![S|\{x_4,x_6\}]\!] = \{s_3,s_6\}$. Since $\triangle\mathcal{L} = -\log 2 + \lambda\log 3 = -0.5245 < 0$, we continue the iterations. At the second iteration, we have $\Pi = \{\{x_3,x_5\},\{x_4,x_6\}\}$ and $\mathcal{Q}^{(2)} = \{\{1,2,7\},\{3,\dots,6\}\}$ after step 11 with the conditional ranges $[\![S|\{x_1,x_2,x_7\}]\!] = \{s_1,s_2,s_6\}$ and $[\![S|\{x_3,\dots,x_6\}]\!] = \{s_3,s_4,s_6\}$, for which $\triangle\mathcal{L} = -0.4605 < 0$. We run the third iteration and have $\mathcal{Q}^{(3)} = \{\{x_1,\dots,x_7\}\}$ returned.*

Let the values of $x\in[\![X]\!]$ be

$$\begin{aligned} &x_1 = 0.2,\ x_2 = 0.1,\ x_3 = 0.4,\ x_4 = 0.3,\\ &x_5 = 0.6,\ x_6 = 1.5,\ x_7 = 1. \end{aligned} \tag{3}$$

and run Algorithm 1 for the utility function $U_2(X;\hat{X}^{(t)}) = -\max_{x,\hat{x}\colon\hat{x}\in[\![\hat{X}^{(t)}|x]\!]} d(x,\hat{x})$ and $\lambda = 2.5$. At the end of 1st iteration, we have $\Pi = \{\{x_2\},\dots,\{x_7\}\}$ and $\mathcal{Q}^{(1)} = \{\{x_1,x_2,x_4\},\{x_3,x_5\},\{x_6,x_7\}\}$ with the conditional ranges $[\![S|\{x_1,x_2,x_4\}]\!] = \{s_1,s_2,s_3\}$, $[\![S|\{x_3,x_5\}]\!] = \{s_3,s_4\}$ and $[\![S|\{x_6,,x_7\}]\!] = \{s_5,s_6\}$. Since $\triangle\mathcal{L} = -\log 2 + \lambda\bar{d}(\{x_6,x_7\}) = -0.3750 < 0$, we continue to iteration $t = 2$ where $\Pi = \{\{x_3,x_5\},\{x_6,x_7\}\}$. But, at step 11, $\mathcal{Q}^{(2)} = \{\{x_1,x_2,x_7\},\{x_3,\dots,x_6\}\}$ but $\triangle\mathcal{L} = -\log 3 + \lambda(\bar{d}(\{x_1,\dots,x_7\}) - \bar{d}(\{x_6,x_7\})) = 0.0758 > 0$. Iteration terminates and the quantization $\mathcal{Q}^{(1)} = \{\{x_1,x_2,x_7\},\{x_3,x_5\},\{x_4,x_6\}\}$ is returned. Algorithm 1 ensures a strict reduction of the Lagrangian function in each iteration. See Fig. 3.

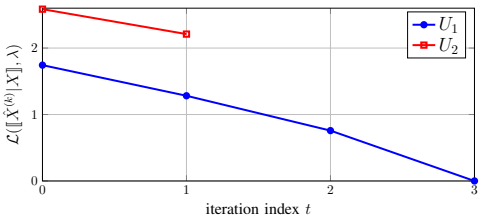


Fig. 3. The convergence performance in terms of $\mathcal{L}([\![\hat{X}^{(t)}|X]\!],\lambda) = -\min_{\mathcal{X}\in\mathcal{Q}^{(t)}}\log|[\![S|\mathcal{X}]\!]| - \lambda U_i(X;\hat{X}^{(t)})$ versus the iteration index t when Algorithm 1 is applied to solve problem (2). For the utility function $U_1(X;\hat{X}^{(t)}) = I_0(S\to\hat{X}^{(t)})$ and $\lambda = 0.3$, the iteration terminates at $t = 3$; For the utility function $U_2(X;\hat{X}^{(t)}) = -\max_{x,\hat{x}\colon\hat{x}\in[\![\hat{X}^{(t)}|X]\!]} d(x,\hat{x})$ and $\lambda = 2.5$, the iteration terminates at $t = 1$. See Example 2. Both plots are strictly decreasing.

B. Minimizing the Maximin Information I_*

Consider the problem

$$\min_{[\![\hat{X}|X]\!]} I_*(S;\hat{X}), \quad \text{s.t. } U_i(X;\hat{X}) \ge \theta. \tag{4}$$

The Lagrangian function of (4) is $\mathcal{L}([\![\hat{X}|X]\!],\lambda) = \log|\mathcal{P}^*_{G_{\hat{X}}}| - \lambda U_i(X;\hat{X})$ for $\lambda\in[0,+\infty)$. Here, $\mathcal{P}^*_{G_{\hat{X}}}$ is the finest decomposition of the undirected uncapacitated graph $G_{\hat{X}} = ([\![\hat{X}]\!],\mathcal{E}_{[\![S|\hat{X}]\!]})$ for the conditional range $[\![S|\hat{X}]\!]$. Based on (1) and Definition 1, we show below how to obtain $\mathcal{P}_{G_{\hat{X}}}$ from $\mathcal{P}_{G_X}$ by a subgraph merging method, the idea of which is then used to propose the agglomerative clustering algorithm for solving problem (4).

1) Determining $\mathcal{P}^*_{G_{\hat{X}}}$: Initiate $\mathcal{P}^*_{G_{\hat{X}}} := \mathcal{P}^*_{G_X}$; For each $\mathcal{X}\in\mathcal{Q} = \{[\![X|\hat{x}]\!]\colon \hat{x}\in[\![\hat{X}]\!]\}$, obtain all the subgraphs in $\mathcal{P}^*_{G_{\hat{X}}}$ that intersect with $\mathcal{X}$ as

$$\mathbb{G}_{\mathcal{X}} := \{\mathcal{G}\in\mathcal{P}^*_{G_{\hat{X}}}\colon \mathcal{G}\cap\mathcal{X}\neq\emptyset\}$$

and update $\mathcal{P}^*_{G_{\hat{X}}}$ by fusing all subgraphs in $\mathbb{G}_{\mathcal{X}}$ to form one subset $\tilde{\mathbb{G}}_{\mathcal{X}} = \sqcup_{\mathcal{G}\in\mathbb{G}_{\mathcal{X}}}\mathcal{G}$, i.e.,

$$\mathcal{P}^*_{G_{\hat{X}}} := (\mathcal{P}^*_{G_{\hat{X}}} \setminus \mathbb{G}_{\mathcal{X}}) \cup \{\tilde{\mathbb{G}}_{\mathcal{X}}\}.$$

The size $|\mathcal{P}^*_{G_{\hat{X}}}|$ is reducing in each iteration and $\mathcal{P}^*_{G_{\hat{X}}}$ is updated to the finest decomposition of $G_{\hat{X}} = (\llbracket\hat{X}\rrbracket, \mathcal{E}_{\llbracket S|\hat{X}\rrbracket})$ at the end such that $I_*(S;\hat{X}) = \log|\mathcal{P}^*_{G_{\hat{X}}}|$. For example, for $\mathcal{P}^*_{G_X} = \{\{x_1,x_2\},\{x_3,x_4\},\{x_5\},\{x_6\},\{x_7\}\}$ in Fig. 2 and the quantization $\mathcal{Q} = \{\{x_1,x_3\},\{x_2,x_5\},\{x_4,x_6\},\{x_7\}\}$, we get the finest decomposition $\mathcal{P}_{G_{\hat{X}}} = \{\{x_1,\ldots,x_6\},\{x_7\}\}$ at the end of the above iterative process (see Fig. 4(c)).

2) Agglomerative Clustering Algorithm: We propose an agglomerative clustering algorithm in Algorithm 2 for problem (4), which can be interpreted as an iterative subgraph merging process. Algorithm 2 starts with $\mathcal{Q}^{(0)} = \{\{x\}\colon x\in\llbracket X\rrbracket\}$. In each iteration t, we have the undirected graph G updated to $G_{\hat{X}^{(t)}} = (\llbracket\hat{X}^{(t)}\rrbracket, \mathcal{E}_{\llbracket S|\hat{X}^{(t)}\rrbracket})$ and the partition $\mathcal{P}^{(t)}$ updated to its finest decomposition $\mathcal{P}^*_{G_{\hat{X}^{(t)}}}$. Here, $\hat{X}^{(t)}$ denotes the released data corresponding to the quantization $\mathcal{Q}^{(t)}$. Therefore, G and $\mathcal{P}^{(0)}$ are initiated as $G_X = (\llbracket X\rrbracket, \mathcal{E}_{\llbracket S|X\rrbracket})$ and $\mathcal{P}^*_{G_X}$, respectively.

Algorithm 2 iteratively merge two subsets $\mathcal{X}_1^*, \mathcal{X}_2^* \in \mathcal{Q}^{(t)}$ that reduce the Lagrangian function $\mathcal{L}(\llbracket\hat{X}^{(t)}|X\rrbracket,\lambda)$ the most (step 4 to step 6). This is done as follows. Consider the Lagrangian function $\mathcal{L}(\llbracket\hat{X}^{(t)}|X\rrbracket,\lambda) = \log|\mathcal{P}^*_{G_{\hat{X}^{(t)}}}| - \lambda U_i(X;\hat{X}^{(t)})$. Since merging subsets of $\mathcal{Q}^{(t)}$ that belong to the same subgraph in $\mathcal{P}^{(t)} = \mathcal{P}^*_{G_{\hat{X}^{(t)}}}$ does not reduce $|\mathcal{P}^*_{G_{\hat{X}^{(t)}}}|$ but only reduce utility $U_i(X;\hat{X}^{(t)})$, we can limit the merging operation to the subsets that belong to disconnected subgraphs in $\mathcal{Q}^{(t)}$. To this end, the set of candidate subsets of $\mathcal{Q}^{(t)}$ for the pairwise merging operation at iteration t is defined as

$$\Psi(\mathcal{Q}^{(t)},\mathcal{P}^{(t)}) = \big\{(\mathcal{X}_1,\mathcal{X}_2)\colon \mathcal{X}_1,\mathcal{X}_2\in\mathcal{Q}^{(t)}, \mathcal{X}_1\neq\mathcal{X}_2, \exists\mathcal{G}_1,\mathcal{G}_2\in\mathcal{P}^{(t)}, \mathcal{G}_1\cap\mathcal{X}_1\neq\emptyset, \mathcal{G}_2\cap\mathcal{X}_2\neq\emptyset, \mathcal{G}_1\neq\mathcal{G}_2\big\}.$$

We determine the steepest decent direction $(\mathcal{X}_1^*,\mathcal{X}_2^*)$ as the maximizer of $\max U_i(X;\hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}})$ over $\Psi(\mathcal{Q}^{(t)},\mathcal{P}^{(t)})$. As explained Section V-A1, for the utility function $U_1(X;\hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}) = I_0(X\to\hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}})$, we choose

$$(\mathcal{X}_1^*,\mathcal{X}_2^*) \in \underset{(\mathcal{X}_1,\mathcal{X}_2)\in\Psi(\mathcal{Q}^{(t)},\mathcal{P}^{(t)})}{\arg\min} \big\{|\mathcal{X}_1|+|\mathcal{X}_2|\big\}. \tag{6}$$

If the minimizer of (6) is not unique, we should choose the pair $(\mathcal{X}_1^*,\mathcal{X}_2^*)$ that connects the largest two subgraphs in $\mathcal{P}^*_{G_{\hat{X}^{(t)}}}$. The purpose is to have a smaller minimum of (6) in the subsequent iterations $t' > t$. See the example below. For the utility function $U_2 = -\max_{x,\hat{x}\colon\hat{x}\in\llbracket\hat{X}|x\rrbracket} d(x,\hat{x})$, we chose

$$(\mathcal{X}_1^*,\mathcal{X}_2^*) \in \underset{(\mathcal{X}_1,\mathcal{X}_2)\in\Psi(\mathcal{Q}^{(t)},\mathcal{P}^{(t)})}{\arg\min} \bar{d}(\mathcal{X}_1\sqcup\mathcal{X}_2).$$

Example 3. *For $\lambda = 0.3$, we apply Algorithm 2 to the uvs S and X with the joint range $\llbracket S,X\rrbracket$ in Example 1 to solve the problem* (4) *for $U_1(X;\hat{X}) = I_0(S\to\hat{X})$. We initiate $\mathcal{Q}^{(0)} = \{\{x_1\},\ldots,\{x_7\}\}$, G as the undirected graph in Fig. 2 and $\mathcal{P}^{(0)} = \{\{x_1,x_2\},\{x_3,x_4\},\{x_5\},\{x_6\},\{x_7\}\}$.*

Algorithm 2: Agglomerative clustering algorithm for solving problem (4)

input : the Lagrangian multiplier $\lambda\in[0,+\infty)$ and the joint range $\llbracket S,X\rrbracket$.
output: the quantization $\mathcal{Q}^{(t)}$, the partition $\mathcal{P}^{(t)}$ that determines the maximin information $I_*(S;\hat{X}^{(t)}) = \log|\mathcal{P}^{(t)}|$ and the graph G that equals $G_{\hat{X}^{(t)}} = (\llbracket\hat{X}^{(t)}\rrbracket, \mathcal{E}_{\llbracket S|\hat{X}^{(t)}\rrbracket})$.

1 initiate the quantization $\mathcal{Q}^{(0)} := \{\{x\}\colon x\in\llbracket X\rrbracket\}$, the graph $G := G_X$ and the finest decomposition $\mathcal{P}^{(0)} := \mathcal{P}^*_{G_X}$;
2 $t := 0$;
3 **while** $|\mathcal{P}^{(t)}| > 1$ **do**
4 Choose a maximizer $(\mathcal{X}_1^*,\mathcal{X}_2^*)$ of
$$\max_{(\mathcal{X}_1,\mathcal{X}_2)\in\Psi(\mathcal{Q}^{(t)},\mathcal{P}^{(t)})} U_i(X;\hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}); \tag{5}$$
5 $\triangle\mathcal{L} :=$
$\log\frac{|\mathcal{P}^{(t)}|-1}{|\mathcal{P}^{(t)}|} - \lambda(U_i(X;\hat{X}^{(t)}_{\{\mathcal{X}_1^*,\mathcal{X}_2^*\}}) - U_i(X;\hat{X}^{(t)}))$;
6 **if** $\triangle\mathcal{L}\geq 0$ **then** break;
7 Merge nodes $\mathcal{X}_1^*$ and $\mathcal{X}_2^*$ in graph G;
8 $\mathcal{Q}^{(t+1)} := \mathcal{Q}^{(t)}_{\{\mathcal{X}_1^*,\mathcal{X}_2^*\}}$;
9 $\mathcal{P}^{(t+1)} := \mathcal{P}^{(t)}_{\{\mathcal{G}_1^*,\mathcal{G}_2^*\}}$, where $\mathcal{G}_1^*,\mathcal{G}_2^*\in\mathcal{P}^{(t)}$ such that $\mathcal{G}_1^*\cap\mathcal{X}_1^*\neq\emptyset$ and $\mathcal{G}_2^*\cap\mathcal{X}_2^*\neq\emptyset$;
10 $t := t+1$;
11 **end**
12 **return** $\mathcal{Q}^{(t)}$, $\mathcal{P}^{(t)}$ and G;

Since $(\{x_1\},\{x_2\})$ and $(\{x_3\},\{x_4\})$ are already connected, candidates for pairwise merging are

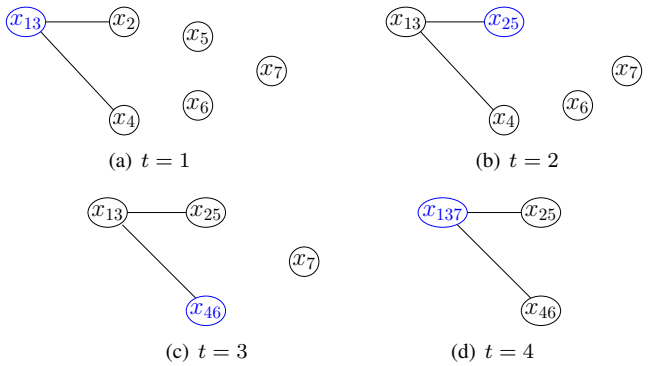
$$\Psi(\mathcal{Q}^{(0)},\mathcal{P}^{(0)}) = \big\{(\{x\},\{x'\})\colon x,x'\in\llbracket X\rrbracket, x\neq x'\big\} \setminus \big\{(\{x_1\},\{x_2\}),(\{x_3\},\{x_4\})\big\}.$$

Note that $|\mathcal{X}_1|+|\mathcal{X}_2| = 2$ for all $(\mathcal{X}_1,\mathcal{X}_2)\in\Psi(\mathcal{Q}^{(0)},\mathcal{P}^{(0)})$. We choose to merge $(\{x_1\},\{x_3\})$ that connect the large two subgraphs $\{x_1,x_2\}$ and $\{x_3,x_4\}$ in the undirected graph in Fig. 2. This results in a strict reduction of the Lagrangian function $\triangle\mathcal{L} = \log_2\frac{4}{5} - \lambda\log_2\frac{2}{1} = -0.0219$.[11] *We get the new quantization $\mathcal{Q}^{(1)} = \{\{x_1,x_3\},\{x_2\},\{x_4\},\ldots,\{x_7\}\}$ and the updated graph G in Fig. 4(a) with the finest decomposition being $\mathcal{P}^{(1)} = \{\{x_1,\ldots,x_4\},\{x_5\},\{x_6\},\{x_7\}\}$. At the 2nd iteration,*

$$\Psi(\mathcal{Q}^{(1)},\mathcal{P}^{(1)}) = \big\{(\mathcal{X}_1,\mathcal{X}_2)\in\mathcal{Q}^{(1)}\colon \mathcal{X}_1\neq\mathcal{X}_2\big\}\setminus \big\{(\{x_1,x_3\},\{x_2\}),(\{x_1,x_3\},\{x_4\}),(\{x_2\},\{x_4\})\big\}.$$

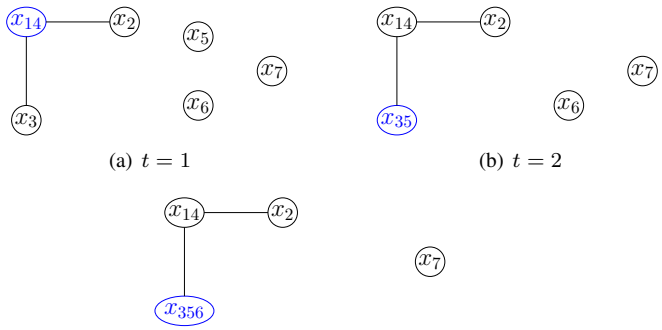
Among the minimizers in $\arg\min\big\{|\mathcal{X}_1|+|\mathcal{X}_2|\colon(\mathcal{X}_1,\mathcal{X}_2)\in\Psi(\mathcal{Q}^{(1)},\mathcal{P}^{(1)})\big\}$, we choose $(\{2\},\{5\})$ that connect the largest two subgraphs $\{1,\ldots,4\}$ and $\{5\}$ and result in $\triangle\mathcal{L} < 0$. The graph G is updated to Fig. 4(b). We have $\mathcal{Q}^{(2)} = \{\{x_1,x_3\},\{x_2,x_5\},\{x_4\},\{x_6\},\{x_7\}\}$ and $\mathcal{P}^{(2)} = \{\{x_1,\ldots,x_5\},\{x_6\},\{x_7\}\}$. Repeating in the same way, we have $\mathcal{Q}^{(3)} = \{\{x_1,x_3\},\{x_2,x_5\},\{x_4,x_6\},\{x_7\}\}$, $\mathcal{P}^{(3)} = \{\{x_1,\ldots,x_6\},\{x_7\}\}$ and the updated graph G in Fig. 4(c) at the end of the 3rd iteration. At the 4th iteration, $\mathcal{Q}^{(4)} =$

[11] One can show that, if $\{x_5\}$ $\{x_6\}$ is merged at the 1st iteration, we have $\min_{(\mathcal{X}_1,\mathcal{X}_2)\in\Psi(\mathcal{Q}^{(2)},\mathcal{P}^{(2)})}\{|\mathcal{X}_1|+|\mathcal{X}_2|\} = 3$ at the 3rd iteration. But in this example, the minimum is 2 when $t=3$.



(a) $t = 1$ (b) $t = 2$ (c) $t = 3$ (d) $t = 4$

Fig. 4. In Example 3, the change of the undirected uncapacitated graph G at the end of each iteration t of Algorithm 2, where $|\mathcal{P}^{(t)}|$, the number of disconnected subgraphs in G, is reduced from $t = 5$ to $t = 1$ and so is $I_*(S; \hat{X}^{(t)})$ from $\log_2 5$ to 0. The subscription of a node in G denotes the original elements in $[\![X]\!]$ that are fused to form this super-node, e.g., x_{13} is the fusion of x_1 and x_3.



(a) $t = 1$ (b) $t = 2$ (c) $t = 3$

Fig. 5. In Example 4, the change of the undirected uncapacitated graph G of the problem at the end of each iteration t of Algorithm 2, where $|\mathcal{P}^{(t)}|$, the number of connected subgraphs is reduced from $|\mathcal{P}^{(0)}| = 5$ (see Fig. 2) to $|\mathcal{P}^{(3)}| = 2$.

$\{\{x_1, x_3, x_7\}, \{x_2, x_5\}, \{x_4, x_6\}\}$, $\mathcal{P}^{(4)} = \{\{x_1, \ldots, x_7\}\}$ *are returned.*

Example 4. *For the values of X in* (3)*, we reset $x_7 = 4$. Consider the utility function $U_2 = -\max_{x,\hat{x}:\, \hat{x}\in[\![\hat{X}|x]\!]} d(x, \hat{x})$. We choose the centroid $\hat{x} = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} x$ for $\mathcal{X} = [\![X|\hat{x}]\!]$ and $d(x, \hat{x}) = \|x - \hat{x}\|_2 = |x - \hat{x}|$. For $\lambda = 0.3$, we apply Algorithm 2 to the uvs S and X with the joint range $[\![S, X]\!]$ in Example 1 to solve the problem* (4)*. We have the graph G updated as in Fig. 5. At the end of the third iteration, $\mathcal{Q}^{(3)} = \{\{x_1, x_4\}, \{x_3, x_5, x_6\}, \{x_2\}, \{x_7\}\}$, $\mathcal{P}^{(3)} = \{\{x_1, \ldots, x_6\}, \{x_7\}\}$ are returned. In Fig. 6, we plot the convergence performance of Algorithm 2 in Examples 3 and 4. It can be seen that Algorithm 2 ensures a strict reduction of the Lagrangian function $L([\![\hat{X}^{(t)}|X]\!], \lambda)$ in each iteration t.*

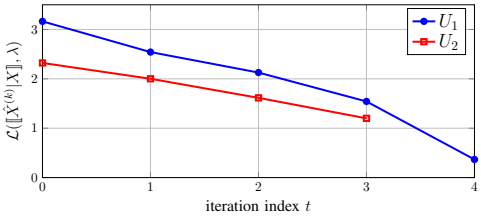


Fig. 6. The convergence performance in terms of $\mathcal{L}([\![\hat{X}^{(t)}|X]\!], \lambda) = \log|\mathcal{P}^{(t)}| - \lambda U_i(X; \hat{X}^{(t)})$ versus the iteration index t when Algorithm 2 is applied to solve problem (4) for $\lambda = 0.3$. For the utility function $U_1(X; \hat{X}^{(t)}) = I_0(S \to \hat{X}^{(t)})$, the iteration terminates at $t = 4$ (see Example 3); For the utility function $U_2(X; \hat{X}^{(t)}) = -\max_{x,\hat{x}:\, \hat{x}\in[\![\hat{X}^{(t)}|x]\!]} d(x, \hat{x})$, the iteration terminates at $t = 3$ (see Example 4). As guaranteed by Algorithm 2, both plots are strictly decreasing.

C. Minimizing $L_0(S \to \hat{X})$ and Restricting $I_*(S; X) = 0$

As discussed in Section III-C, we have $L_0(S \to \hat{X}) - I_*(S; \hat{X}) \geq 0$, where $I_*(S; \hat{X})$ can be considered as the most bits in the leakage $L_0(S \to \hat{X})$ that can be estimated/distinguished by an adversary without error. Also, due to the fact that the privacy leakage $L_0(S \to \hat{X})$ could be nonzero when $I_*(S; \hat{X}) = 0$, it is worth considering the problem of how to find a data release scheme that ensures the non-stochastic indistinguishability

$$\min_{[\![\hat{X}|X]\!]} L_0(S \to \hat{X}), \quad \text{s.t. } I_*(S; X) = 0, U_i(X; \hat{X}) \geq \theta. \quad (7)$$

In (7), we assume that $\theta \geq \min_{[\![\hat{X}|X]\!]:\, I_*(S;\hat{X})=0} U_i(X; \hat{X})$.[12] The Lagrangian function is

$$\mathcal{L}([\![S|X]\!], \lambda', \lambda) = - \min_{\mathcal{X}\in\mathcal{Q}} \log|[\![S|\mathcal{X}]\!]| + \lambda' \log|\mathcal{P}_{G_{\hat{X}}}| - \lambda U_i(X; \hat{X}) \quad (8)$$

for $\lambda', \lambda \geq 0$. The same as Algorithm 2, the steepest decent direction $(\mathcal{X}_1^*, \mathcal{X}_2^*)$ can still be searched over the set $\Psi(\mathcal{P}^{(t)}, \mathcal{Q}^{(t)})$, but the iteration should repeat until $|\mathcal{P}^{(t)}| = 1$. The resulting algorithm is the same as Algorithm 2 except that steps 4 to 6 are replaced by

$$(\mathcal{X}_1^*, \mathcal{X}_2^*) := \operatorname*{arg\,min}_{(\mathcal{X}_1,\mathcal{X}_2)\in\Psi(\mathcal{P}^{(t)},\mathcal{Q}^{(t)})} \triangle\mathcal{L}, \quad (9)$$

where

$$\triangle\mathcal{L} = - \min_{\mathcal{X}\in\mathcal{Q}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}}} \log|[\![S|\mathcal{X}]\!]| + \min_{\mathcal{X}\in\mathcal{Q}^{(t)}} \log|[\![S|\mathcal{X}]\!]| + \lambda(U_i(X; \hat{X}^{(t)}) - U_i(X; \hat{X}^{(t)}_{\{\mathcal{X}_1,\mathcal{X}_2\}})).$$

If the minimizers of (9) is not unique, we choose $(\mathcal{X}_1^*, \mathcal{X}_2^*)$ with the minimum $|[\![S|\mathcal{X}_1^*]\!]|$ and $|[\![S|\mathcal{X}_2^*]\!]|$. Note, the above method ensures a strict reduction of $I_*(S; \hat{X}^{(t)})$ for each iteration t, which is equivalent to the case $\lambda' \gg \lambda$.

Example 5. *We apply the above method to the uvs S and X in Example 1 for the utility function $U_1(X; \hat{X}) = I_0(S \to \hat{X}$ and $\lambda = 0.3$. Starting with $\mathcal{Q}^{(0)} = \{\{x_1\}, \ldots, \{x_7\}\}$ and*

[12]This is to ensure a nonempty feasible region for (7). The proposed clustering algorithm based on the Lagrangian function (8) does not require the value of $\min_{[\![\hat{X}|X]\!]:\, I_*(S;\hat{X})=0} U_i(X; \hat{X})$ in advance, since the returned solution $\hat{X}$ for any given λ is always feasible.

$\mathcal{P}^{(0)} = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5\}, \{x_6\}, \{x_7\}\}$*, at the first iteration, we have* $\arg\min\{\triangle L\colon (\mathcal{X}_1, \mathcal{X}_2) \in \Psi(\mathcal{P}^{(0)}, \mathcal{Q}^{(0)})\} = \Psi(\mathcal{P}^{(0)}, \mathcal{Q}^{(0)})$*, where we choose* $(\mathcal{X}_1^*, \mathcal{X}_2^*) = (\{2\}, \{3\})$ *such that* $|[\![S|\{2\}]\!]| = |[\![S|\{3\}]\!]| = 2$ *are the largest subgraphs. We then have* $\mathcal{Q}^{(1)} = \{\{x_1\}, \{x_2, x_3\}, \{x_4\}, \ldots, \{x_7\}\}$ *and* $\mathcal{P}^{(1)} = \{\{x_1, \ldots, x_4\}, \{x_5\}, \{x_6\}, \{x_7\}\}$*. At the second iteration, we choose* $(\mathcal{X}_1^*, \mathcal{X}_2^*) = (\{x_4\}, \{x_5\})$ *and have* $\mathcal{Q}^{(2)} = \{\{x_1\}, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6\}, \{x_7\}\}$ *and* $\mathcal{P}^{(2)} = \{\{x_1, \ldots, x_5\}, \{x_6\}, \{x_7\}\}$*. At the third iteration, we have* $(\mathcal{X}_1^*, \mathcal{X}_2^*) = (\{x_6\}, \{x_7\})$ *so that* $\mathcal{Q}^{(3)} = \{\{x_1\}, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7\}\}$ *and* $\mathcal{P}^{(3)} = \{\{x_1, \ldots, x_5\}, \{x_6, x_7\}\}$*. At the fourth iteration, there are only two subgraphs* $\{x_1, \ldots, x_5\}$ *and* $\{x_6, x_7\}$ *in* $G_{\hat{X}^{(t)}}$*. We have* $(\mathcal{X}_1^*, \mathcal{X}_2^*) = (\{x_1\}, \{x_6, x_7\})$ *so that* $|\mathcal{P}^{(4)}| = |\{\{x_1, \ldots, x_7\}\}| = 1$ *finally. The quantization method* $\mathcal{Q}^{(4)} = \{\{x_1, x_6, x_7\}, \{x_2, x_3\}, \{x_4, x_5\}\}$ *is returned.*

Note, when the sensitive data S is also observable or available, the data curator can sanitize the data by quantizing the joint range $[\![S, X]\!]$. It can be modeled by Markov chain $S - Y - \hat{Y}$, where $Y = (S, X)$ denotes the public data and $\hat{Y}$ is its sanitization. The purpose is to search a privacy-preserving quantization function $f\colon [\![Y]\!] \mapsto [\![\hat{Y}]\!]$ that determines $[\![\hat{Y}|Y]\!]$, for which, the proposed algorithms in this section also apply by replacing X and $\hat{X}$ with Y and $\hat{Y}$, respectively.

It is also easy to extend to a system model where the adversary has side information A. With the range of the sensitive data S conditioned on A before and after the data release $\hat{X}$ denoted by $[\![S|A]\!]$ and $[\![S|\hat{X}, A]\!]$, respectively, the privacy leakage $L_0(S \to \hat{X}|A) = \max_{(\hat{x},a)\in[\![\hat{X},A]\!]} \log \frac{|[\![S|a]\!]|}{|[\![S|\hat{x},a]\!]|}$ and the maximin information $I_*(S; \hat{X}, A) = \log |\mathcal{P}^*_{[\![S|\hat{X},A]\!]}|$ can be used as the privacy measures. Here, $\mathcal{P}^*_{[\![S|\hat{X},A]\!]}$ refers to the finest decomposition of the undirected graph $G_{X,A} = ([\![X, A]\!], \mathcal{E}_{[\![S|X,A]\!]})$. In this case, we are instead considering the problem $\min_{[\![\hat{X}|X]\!]} L_0(S \to \hat{X}|A)$ or $I_*(S; \hat{X}, A)$, s.t. $U_i(X; \hat{X}) \geq \theta$, the solution to which can also be searched by Algorithm 1 or 2.

VI. Notes on Stochastic Information-theoretic Privacy and Differential Privacy

The studies on privacy in traditional (i.e., stochastic) information theory [7], [8] consider random variables (rvs) S and X, and assumes that an adversary infers statistical knowledge on S via X. By considering the correlation, the joint probability $p(s, x), \forall s, x$, between the private and public data, the problem is to design a private encoder/channel $p(\hat{x}|x), \forall x, \hat{x}$ that minimizes privacy leakage, but maintains some level of data utility. This framework is also captured by the Markov chain $S - X - \hat{X}$, where $p(\hat{x}|x)$ is called the *privacy funnel* [34]. The design of the privacy funnel $p(\hat{x}|x)$ is shown to be an opposite optimization problem to the *information bottleneck*[13] [35], [36]. The idea of Algorithms 1 and 2 in this paper is analogous to the agglomerative pairwise merge algorithms in [34], [36]. Similar to $L_0(S \to \hat{X})$, privacy in information theory is measured in terms of the logarithm of the fraction between the prior and posterior statistical uncertainty on S. For example, the average inference loss defined as $H(S) - H(S|\hat{X})$, or the mutual information $I(S; \hat{X})$, is extended to the worst case $H(S) - \max_{\hat{x}} H(S|\hat{x})$ [7] and the α-leakage [9], [37], where the latter is a tunable measure of the mutual information.

A. Privacy Leakage L_0 and Stochastic Privacy Measures

Let S be a uniformly distributed rv and consider the maximization of Rényi divergence $D_\alpha(p(s|\hat{x})\|p(s)) = \frac{1}{\alpha-1}\log\sum_{\hat{x}} \frac{p^\alpha(s|\hat{x})}{p^{\alpha-1}(s)}$ in the order of $\alpha = 0$ over all $\hat{x} \in [\![\hat{X}]\!]$. We have

$$\begin{aligned}\max_{\hat{x}\in[\![\hat{X}]\!]} D_0(p(s|\hat{x})\|p(s)) &= \max_{\hat{x}\in[\![\hat{X}]\!]} -\log \sum_{s\colon p(s|\hat{x})>0} p(s) \\ &= \max_{\hat{x}\in[\![\hat{X}]\!]} -\log \sum_{s\colon p(s|\hat{x})>0} \frac{1}{|[\![S]\!]|} \\ &= \max_{\hat{x}\in[\![\hat{X}]\!]} \log \frac{|[\![S]\!]|}{|[\![S|\hat{x}]\!]|} \\ &= L_0(S \to \hat{X}).\end{aligned} \tag{10}$$

When the order $\alpha \to \infty$,

$$\begin{aligned}\max_{\hat{x}\in[\![\hat{X}]\!]} D_\infty(p(s|\hat{x})\|p(s)) &= \max_{(s,\hat{x})\in[\![S,\hat{X}]\!]} \log \frac{p(s|\hat{x})}{p(s)} \\ &= D_\infty(p(s,\hat{x})\|p(s)p(\hat{x})) = \mathcal{L}^r(S \to \hat{X}).\end{aligned}$$

Here, $\mathcal{L}^r(S \to \hat{X})$ is called the *maximal realizable leakage* in [38, Definition 8]. It is also shown in [7, Definition 6 and Theorem 3] that if

$$\max_{(s,\hat{x})\in[\![S,\hat{X}]\!]} \Big|\log \frac{p(s|\hat{x})}{p(s)}\Big| = \epsilon, \tag{11}$$

the randomization $p(\hat{x}|s)$ is at least 2ϵ-locally differential private (2ϵ-LDP)[14].

In [38, Theorem 1], the stochastic maximal leakage $\mathcal{L}(S \to \hat{X})$ is defined and shown to be equivalent to the Sibson mutual information $I_\infty(S; \hat{X})$, which is upper bounded as $\mathcal{L}(S \to \hat{X}) \leq \mathcal{L}^r(S \to \hat{X})$ [38, Section VI-B]. Meanwhile, a more general measure, the maximal α-leakage $\mathcal{L}_\alpha^{\max}(S \to \hat{X})$, was proposed in [9, Definition 6] based on the Arimoto mutual information, which reduces to the maximal leakage $\mathcal{L}_\infty^{\max}(S \to \hat{X}) = \mathcal{L}(S \to \hat{X})$ when the order $\alpha \to \infty$ [9, Theorem 2]. It is also shown in [9, Theorem 2] that when $\alpha = 1$, the maximal α-leakage $\mathcal{L}_\alpha^{\max}(S \to \hat{X})$ reduces to Shannon mutual information $I(S; \hat{X})$, which was used as the privacy measure, called *synergistic disclosure capacity*, in [41, Definition 2]. The mutual information $I(S; \hat{X})$ is equivalent to the f-information $I_f(S; \hat{X})$, e.g., used in [42], in the case of $f(t) = t \log t$.

[13]For the utility function $\mathbb{E}_{x,\hat{x}}[d(x, \hat{x})]$, the information bottleneck generalizes the rate-distortion problem in information theory [35].

[14]A data release scheme $p(x|s)$ is ϵ-LDP if $p(x|s)/p(x|s') \leq e^\epsilon, \forall x, s, s'$ [39]. Note that LDP is equivalent to DP [4] without the constraint on the Hamming distance $d_H(s, s') \leq 1$. This definition is adopted in information theory, e.g., [40].

B. Maximin Information I_* and Gács-Körner Common Information

The maximin information $I_*(S;\hat{X})$ is related to the Gács-Körner common information [20]:

$$K(S;\hat{X}) = -\sum_{\mathcal{S}\in[\![S|\hat{X}]\!]_*} p(\mathcal{S})\log p(\mathcal{S}),$$

where $p(\mathcal{S}) = \sum_{s\in\mathcal{S},\hat{x}\in[\![\hat{X}|s]\!]} p(s,\hat{x})$ and $[\![S,\hat{X}]\!] = \{(s,\hat{x})\colon p(s,\hat{x}) \neq 0\}$. Note that $I_*(S;\hat{X}) = \max_{p(s,\hat{x}),\forall s,\hat{x}} K(S;\hat{X})$ and therefore $K(S;\hat{X})$ denotes the number of bits in S that can be perfectly received by the adversary via the privacy funnel $p(\hat{x}|x)$. The sample complexity[15] for empirically learning the value $K(S;\hat{X})$ [43, Proposition 2] also applies to $I_*(S;\hat{X})$.

Similar to $I_*(S;\hat{X}) \le L_0(S\to\hat{X})$, we have $K(S;\hat{X}) \le I(S;\hat{X})$ [20], [44]. This means that, as long as $I(S;\hat{X}) > 0$, minimizing $I(S;\hat{X})$, or any other stronger privacy leakage, e.g., the maximal leakage [38], does not necessarily ensure $K(S;\hat{X}) = 0$. On the other hand, the case $I_*(S;\hat{X}) > 0$ indicates that the LDP [4] is unattainable. This is because there exists an x such that $\log(p(x|s)/p(x|s')) \to \infty$ for some pair of s and s'.

Note that, in stochastic information theory, *perfect privacy* [45], [46] refers to the independence $S \perp \hat{X}$, i.e., $p(s|\hat{x}) = p(s), \forall\hat{x}$. In this case, $I(S;\hat{X}) = 0$ so that $K(S;\hat{X}) = 0$, necessarily. Although the condition $K(S;\hat{X}) = 0$ alone does not guarantee perfect privacy[16], the value of the Gács-Körner common information indicates an upper bound on the utility of the released data in the case of the perfect privacy (see [8]). The case $S \perp \hat{X}$ corresponds to non-stochastic independence $[\![S|\hat{x}]\!] = [\![S]\!], \forall\hat{x}$, where $I_0(S\to\hat{X}) = L_0(S\to\hat{X}) = 0$ and $I_*(S;\hat{X}) = 0$. But, since $I_*(S;\hat{X}) \le L_0(S\to\hat{X})$, $I_*(S;\hat{X}) = 0$ does not guarantee $[\![S|\hat{x}]\!] = [\![S]\!], \forall\hat{x}$, either. See also [15, Fig. 5]. Therefore, the solution to problem (7) is not necessarily a data release scheme with perfect non-stochastic privacy.

VII. Experiments

In this section, we run experiments to show the performance of Algorithm 1 and 2 proposed in Section V. First, we show the Pareto frontier of Example 1 to explain how to search the solution to problems (2), (4) and (7). Then, we present both the convergence performance and Pareto frontier of Algorithms 1 and 2 when they are applied to the real-world dataset.

A. Pareto Frontier

The Lagrangian function $\mathcal{L}([\![\hat{X}|X]\!],\lambda)$ can be considered as the weighted-sum of two objective functions: the privacy leakage $I_*(S;\hat{X})$, or $L_0(S\to\hat{X})$, and the utility loss $-U_i(X;\hat{X})$. They form the PUT: minimizing one necessarily increases the other. In this case, the minimizer of the problem $\min_{[\![\hat{X}|X]\!]}\mathcal{L}([\![\hat{X}|X]\!],\lambda)$ for each value of λ produces a *Pareto optimal* pair $\{I_*(S;\hat{X}), -U_i(X;\hat{X})\}$, or $\{L_0(S\to\hat{X}), -U_i(X;\hat{X})\}$, indicating the smallest privacy leakage $I_*(S;\hat{X})$ or $L_0(S\to\hat{X})$ that can be attained by the design of the mapping $[\![\hat{X}|X]\!]$ to maintain certain level θ of $U_i(X;\hat{X})$ and vice versa [47]. The Pareto optimal pair for all value of λ forms the *Pareto frontier*.

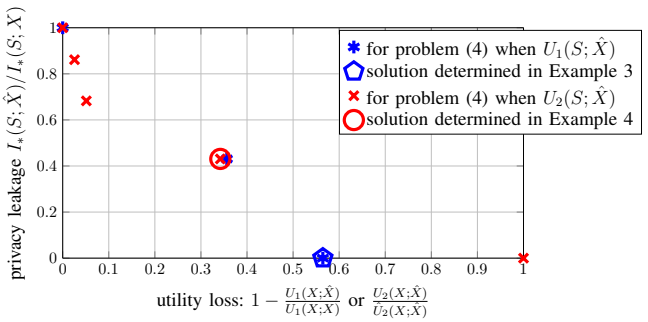


Fig. 7. Pareto frontier for problem (4) for both utility functions $U_1(S;\hat{X}) = I_0(X\to\hat{X})$ and $U_2(S;\hat{X}) = -\max_{x,\hat{x}\colon\hat{x}\in[\![\hat{X}|x]\!]} d(x,\hat{x})$. Note, both privacy and utility measures are normalized. For $U_1(S;\hat{X})$, the horizontal axis is $1-\frac{U_1(X;\hat{X})}{U_1(X;X)} = 1-\frac{I_0(X\to\hat{X})}{H_0(X)}$; For $U_1(S;\hat{X})$, the horizontal axis is $\frac{U_2(X;\hat{X})}{U_2(X;\hat{X})}$, where $\hat{U}_2(X;\hat{X})$ is the minimum value of $U_2(X;\hat{X})$ over all points in the Pareto frontier.

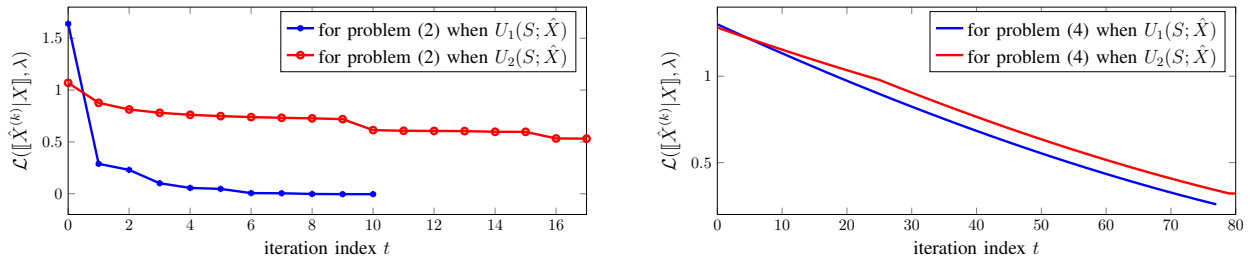
Fig. 7 shows the examples of the Pareto frontier for the problem (4), where we can see the solutions obtained in Examples 3 and 4. Taking the solution in Example 3 for example. It corresponds to the pair: $I_*(S;\hat{X}) = 0$ and $I_0(X;\hat{X}) = \log 7/3$, which states that, to maintain the utility no less than $\log 7/3$, we can reduce the privacy leakage $I_*(S;\hat{X})$ to zero. When applying the algorithms proposed in this paper to the real-world dataset, enumerate λ to obtain the Pareto frontier. For given utility threshold θ, search the Pareto frontier for the minimal privacy leakage and the corresponding solution $[\![\hat{X}|X]\!]$.

B. Hungarian Heart Disease Dataset

This experiment is based on the heart disease data set created by the Hungarian Institute of Cardiology, Budapest, in the UCI machine learning repository [48]. It records 293 patients' data of 76 attributes for the purpose of identifying the presence of heart disease. We extract the column 'age' as the sensitive data S and 'serum cholesterol (mg/dl)' as the public data X. The marginal range $[\![S]\!]$ contains 38 distinct values of age and $[\![X]\!]$ contains 154 cholesterol levels. The joint range $[\![S,X]\!]$ contains 281 distinct values of (s,x), among which 269 values appear only once in the whole dataset. Note that this is a typical example that the size of released dataset may not be large enough to reveal the true statistics of the data. In this case, each data record often releases a new value of (s,x) and the adversary only infers the supports or cardinalities of the alphabets $\mathcal{S}$ and $\mathcal{X}$, not the probability or data statistics.

[15]The sample complexity derived in [43, Proposition 2] applies to learning the Hirschfeld-Gebelein-Rényi (HGR) maximal correlation function, which determines $K(S;\hat{X})$ [33, Corollary 3, Remark 6] and therefore applies to $I_*(S;\hat{X})$.

[16]The interpretation is that even if the DP is attainable, the information leakage could be nonzero.



(a) $\lambda = 2.3$ for $U_1(X;\hat{X}^{(t)})$; $\lambda = 2$ for $U_2(X;\hat{X}^{(t)})$ (b) $\lambda = 1.3$ for $U_1(X;\hat{X}^{(t)})$; $\lambda = 0.03$ for $U_2(X;\hat{X}^{(t)})$

Fig. 8. The convergence performance in terms of $\mathcal{L}([\![\hat{X}^{(t)}|X]\!],\lambda)$ versus the iteration index t when Algorithms 1 and 2 are applied to the Hungarian heart disease dataset in [48] for solving problems (2) and (4), respectively. Here, $U_1(X;\hat{X}^{(t)}) = I_0(S \to \hat{X}^{(t)})$ and $U_2(X;\hat{X}^{(t)}) = -\max_{x,\hat{x}:\ \hat{x}\in[\![\hat{X}^{(t)}|x]\!]} d(x,\hat{x})$.

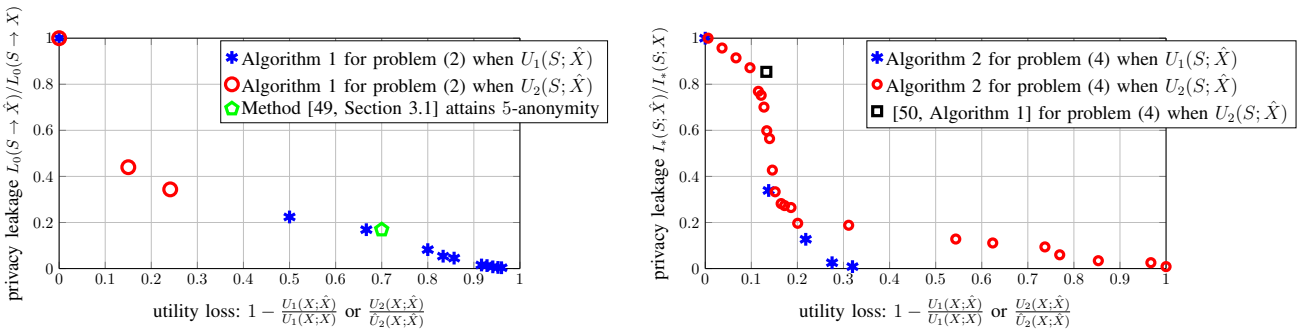


Fig. 9. Pareto frontiers for the problems (2) and (4) for both utility functions $U_1(S;\hat{X}) = I_0(X \to \hat{X})$ and $U_2(S;\hat{X}) = -\max_{x,\hat{x}:\ \hat{x}\in[\![\hat{X}|x]\!]} d(x,\hat{x})$, obtained by applying Algorithms 1 and 2 to the Hungarian heart disease dataset in [48]. They are compared to the solution search by the generalization-and-suppression algorithm for attaining the 5-anonymity in [49, Section 3.1] and the submodularity-based agglomerative clustering algorithm in [50, Algorithm 1].

Therefore, the variables S and X can be treated as uncertain variables and the information can be directly obtained from the ranges of S and X by the non-stochastic information measures in Section III.

Fig. 8 shows that Algorithms 1 and 2 converge to the locally optimal solutions to problems (2) and (4), respectively. Fig. 9 shows the Pareto frontier searched by Algorithms 1 and 2 by enumerating the value of Lagrangian multiplier λ.

Because of the equivalence of k-anonymity and L_0 in Lemma 1, we also run the generalization-and-suppression method proposed [49, Section 3.1] for attaining the k-anonymity for $k = 5$. The idea is to search for all x associated with less than 5 values of s and merge them to a new but less specific value, which is exactly the same idea as the agglomerative clustering algorithms proposed in this paper. The only difference is that the generalization-and-suppression method does not guarantee a certain level of utility, which means it may sacrifice too much data utility to attain the k-anonymity. As shown in Fig. 9, the solution determined by Algorithm 1 outperforms the generalization-and-suppression method proposed in [49, Section 3.1] in that it incurs a lower utility loss, or maintains a higher level of usefulness in the released dataset.[17]

For problem (4), we run the submodularity-based agglomerative clustering algorithm proposed in [50, Algorithm 1] and compare with Algorithm 2 in the Pareto frontier in Fig. 9. It can be seen that Algorithm 2 is able to search a solution that *Pareto dominates* [50, Algorithm 1]: One can find several points determined by Algorithm 2 with both privacy leakage and utility loss strictly less than the solution searched by [50, Algorithm 1]. This is because the problem under concern is not submodular[18] and therefore submodularity-based algorithms may not be superior to greedy ones.

VIII. Conclusion

We studied the problem of how to quantize the public data X into $\hat{X}$ to minimize the non-stochastic information leakage of the private data S but guarantee a certain level of the data utility. For the privacy measures $L_0(S \to \hat{X})$ and $I_*(S;X)$, two agglomerative clustering algorithms were proposed, respectively, both of which generate a solution $[\![\hat{X}|X]\!]$ by recursively merging two elements in $[\![X]\!]$ that strictly reduces the Lagrangian function. We showed that $L_0(S \to \hat{X})$ measures the maximum (worst-case) posterior range/uncertainty reduction on S at the adversary side and $I_*(S;\hat{X})$ denotes how distinguishable S is by observing $\hat{X}$. We then applied the clustering algorithm for minimizing $I_*(S;X)$ to search for a $\hat{X}$ that guarantees non-stochastic indistinguishability $I_*(S;\hat{X}) = 0$ but minimizes $L_0(S \to \hat{X})$

[17]It can be seen by comparing to the blue star searched by Algorithm 1 on the left-hand side of the green pentagon determined by the generalization-and-suppression method proposed in [49, Section 3.1].

[18]The validity and efficiency of the agglomerative clustering algorithm in [50, Algorithm 1] is based on the submodularity of the Lagrangian function. However, in this paper, the Lagrangian function $\mathcal{L}([\![\hat{X}^{(t)}|X]\!],\lambda)$ for both privacy measures I_* and L_0 does not exhibit submodularity or supermodularity. One can prove by counter-examples based on the definition of submodularity in [28], [51].

subject to some utility constraint. It is shown that the value of $I_*(S;\hat{X})$ is equal to the maximum number of disconnected subgraphs in the confusability graph, which can be determined by the min-cut or optimal network attack algorithms. This gives a visualization of the clustering algorithm for minimizing $I_*(S;X)$ (Algorithm 2) as a subgraph merging process.

There are some aspects in this paper that can be further explored. First, the proposed greedy clustering algorithms only converge to a local optimum and have the complexity $O(|[\![X]\!]|^3)$. It is worth discussing how to improve the accuracy and efficiency. Second, based on Section VI, we can combine the results in stochastic and non-stochastic information-theoretic privacy in real applications, e.g., adopt non-stochastic measure $L_0(S \to \hat{X})$ at the beginning of the data release and then switch[19] to stochastic measure $\max_{\hat{x}\in[\![\hat{X}]\!]} D_\alpha(p(s|\hat{x})\|p(s))$, which is done by increasing α from 0 based on (10), or start with $I_*(S;\hat{X})$ and switch to $K(S;\hat{X})$. In addition, the non-stochastic privacy measures can also be utilized to design stochastic privacy-preserving mechanisms: determining the optimal transition probability as well as the optimal alphabet size.[20] Third, the same as the Gács-Körner common information, the maximin information is zero in most cases. Therefore, it is of interest to consider other non-stochastic measure of the common information, e.g., to propose a non-stochastic version of the Wyner's common information in [55], and explore its role in perfect privacy.

Appendix A
Proof of Proposition 1

For the finest $[\![S|X]\!]$-overlap partition $\mathcal{P}^*_{[\![S|X]\!]}$ and the maximin information $I_*(S;X) = \log|\mathcal{P}^*_{[\![S|X]\!]}|$, we have the Hartley entropy of S being

$$\begin{aligned} H_0(S) &= \log|[\![S]\!]| \\ &= \log \sum_{\mathcal{S}\in\mathcal{P}^*_{[\![S|X]\!]}} |\mathcal{S}| \\ &= I_*(S;X) + \log\Big(\frac{1}{|\mathcal{P}^*_{[\![S|X]\!]}|}\sum_{\mathcal{S}\in\mathcal{P}^*_{[\![S|X]\!]}}|\mathcal{S}|\Big) \qquad (12)\\ &\geq I_*(S;X) + \log\min_{\mathcal{S}\in\mathcal{P}^*_{[\![S|X]\!]}}|\mathcal{S}| \\ &\geq I_*(S;X) + \log\min_{x\in[\![X]\!]}|[\![S|x]\!]|. \end{aligned}$$

This proves $I_*(S;X) \leq \log\frac{|[\![S]\!]|}{\min_{x\in[\![X]\!]}|[\![S|x]\!]|} = \max_{x\in[\![X]\!]}\log\frac{|[\![S]\!]|}{|[\![S|x]\!]|} = L_0(S\to X)$. Note that the inequality $K(S;X) \leq I(S;X)$ [20], [44] (see Section VI) describing the relation between the Gács-Körner common information $K(S;X)$ and Shannon mutual information $I(S;X)$ in stochastic information theory can be proved in the same way as (12) based on the Jensen's inequality. ■

Appendix B
Proof of Lemma 2

For $\mathcal{X}\subseteq[\![X]\!]$, denote $[\![S|\mathcal{X}]\!] = \bigcup_{x\in\mathcal{X}}[\![S|X=x]\!]$. Since there is no edge connecting any distinct subgraphs $\mathcal{X},\mathcal{X}'\in\mathcal{P}_{[\![X]\!]}$, i.e., $[\![S|\mathcal{X}]\!]\cap[\![S|\mathcal{X}']\!]=\emptyset$, and $\bigcup_{\mathcal{X}\in\mathcal{P}_{[\![X]\!]}}[\![S|\mathcal{X}]\!] = \bigcup_{x\in[\![X]\!]}[\![S|X=x]\!] = [\![S]\!]$. So, $\{[\![S|\mathcal{X}]\!]:\mathcal{X}\in\mathcal{P}_{G_X}\}$ for every $\mathcal{P}_{G_X}$ is a partition of $[\![S]\!]$. For any two points $x\in\mathcal{X}$ and $x'\in\mathcal{X}'$ such that $\mathcal{X},\mathcal{X}'$ are two distinct subgraphs in $\mathcal{P}_{G_X}$, there does not exist a path, denoted by an ordered sequence $(x_1,\ldots,x_n)$, such that $x_1=x$ and $x_n=x'$ and $(x_i,x_{i+1})\in\mathcal{E}_{[\![S|X]\!]}$, i.e., $[\![S|X=x_i]\!]\cap[\![S|X=x_{i+1}]\!]\neq\emptyset$. This means any $s\in[\![S|X=x]\!]\subseteq[\![S|\mathcal{X}]\!]$ and $s'\in[\![S|X=x']\!]\subseteq[\![S|\mathcal{X}']\!]$ are overlap isolated. Therefore, $\mathcal{P}_{G_X}$ is a $[\![S|X]\!]$-overlap isolated partition In the finest decomposition $\mathcal{P}^*_{G_X}$, for each $\mathcal{X}\in\mathcal{P}^*_{G_X}$, every pair $x,x'\in\mathcal{X}$ are connected, i.e., there exists a path $(x_1=x,\ldots,x_n=x')$ such that $(x_i,x_{i+1})\in\mathcal{E}_{[\![S|X]\!]}$ or $[\![S|X=x_i]\!]\cap[\![S|X=x_{i+1}]\!]\neq\emptyset$. Then, we have $s\leftrightsquigarrow s'$ for any two $s,s'\in[\![S|\mathcal{X}]\!]$. So, $\{[\![S|\mathcal{X}]\!]:\mathcal{X}\in\mathcal{P}^*_{G_X}\}$ is $[\![S|X]\!]$-overlap partition and also the finest one since it can not be further decomposed. ■

[19]For example, change to a stochastic measure when the amount of the released data is enough (say, 5000 tabular records). In this case, it is also crucial to determine the optimal switching time.

[20]Most stochastic privacy-preserving solutions only determine the optimal transition probability $p(\hat{x}|x)$ for fixed size of $[\![\hat{X}]\!]$, e.g., the variational method in [35] for information bottleneck, the Blahut–Arimoto algorithm in [52] for rate-distortion. An iterative approach can be proposed by combining these methods with Algorithms 1 and 2 to first determine the optimal alphabet $[\![\hat{X}]\!]$, then the optimal transition probability $p(\hat{x}|x)$ (see also the deterministic annealing method in [53]). A related study can be found in [54], which shows how, in information bottleneck, the error of empirical mutual information is upper bounded by the alphabet size.

References

[1] J. Lane, V. Stodden, S. Bender, and H. Nissenbaum, *Privacy, big data, and the public good: Frameworks for engagement*. New York, NY: Cambridge University Press, 2014.

[2] N. Li, W. Qardaji, D. Su, Y. Wu, and W. Yang, "Membership privacy: a unifying framework for privacy definitions," in *Proc. of 2013 ACM SIGSAC conf. Comput. Commun. Secur.*, Berlin, Germany, 2013, pp. 889–900.

[3] D. J. Glancy, "The invention of the right to privacy," *Arizona Law Rev.*, vol. 21, no. 21, pp. 1–39, 1979.

[4] C. Dwork, "Differential privacy," *Encycl. Cryptography Secur.*, pp. 338–340, 2011.

[5] K. Chaudhuri and C. Monteleoni, "Privacy-preserving logistic regression," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, B.C., 2009, pp. 289–296.

[6] A. Friedman and A. Schuster, "Data mining with differential privacy," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Min.*, Washington, DC, 2010, pp. 493–502.

[7] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in *Proc. 50th Ann. Allerton Conf. Comm. Control Comput.*, Monticello, IL, 2012, pp. 1401–1408.

[8] S. Asoodeh, F. Alajaji, and T. Linder, "Notes on information-theoretic privacy," in *Proc. 52th Ann. Allerton Conf. Comm. Control Comput.*, Monticello, IL, 2014, pp. 1272–1278.

[9] J. Liao, O. Kosut, L. Sankar, and F. d. P. Calmon, "Tunable measures for information leakage and applications to privacy-utility tradeoffs," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8043–8066, 2019.

[10] C. Dwork and V. Feldman, "Privacy-preserving prediction," *arXiv preprint*, 2018. [Online]. Available: https://arxiv.org/abs/1803.10266

[11] A. N. Kolmogorov and V. M. Tikhomirov, "ε-entropy and ε-capacity of sets in function spaces," *Uspekhi Matematicheskikh Nauk*, vol. 14, no. 2, pp. 3–86, 1959.

[12] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 10, no. 05, pp. 557–570, 2002.

[13] F. Farokhi, "Development and analysis of deterministic privacy-preserving policies using non- stochastic information theory," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 10, pp. 2567–2576, Oct. 2019.

[14] F. Farokhi, "Non-stochastic hypothesis testing with application to privacy against hypothesis-testing adversary," *arXiv preprint*, 2019. [Online]. Available: https://arxiv.org/abs/1904.07377

[15] G. N. Nair, "A nonstochastic information theory for communication and state estimation," *IEEE Trans. Autom. Control*, vol. 58, no. 6, pp. 1497–1510, Jun. 2013.
[16] A. Saberi, F. Farokhi, and G. N. Nair, "State estimation over worst-case erasure and symmetric channels with memory," *arXiv preprint*, 2019. [Online]. Available: https://arxiv.org/abs/1902.00726
[17] P. Duan, F. Yang, S. L. Shah, and T. Chen, "Transfer zero-entropy and its application for capturing cause and effect relationship between variables," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 3, pp. 855–867, May 2015.
[18] M. Wiese, K. H. Johansson, T. J. Oechtering, P. Papadimitratos, H. Sandberg, and M. Skoglund, "Uncertain wiretap channels and secure estimation," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, 2016, pp. 2004–2008.
[19] M. J. Erickson, *Introduction to combinatorics*, 2nd ed. New Jersey: John Wiley & Sons, 2013.
[20] P. Gács and J. Körner, "Common information is far less than mutual information," *Probl. Control Inf. Theory*, vol. 2, no. 2, pp. 149–162, 1973.
[21] R. V. Hartley, "Transmission of information 1," *Bell Syst. Tech. J.*, vol. 7, no. 3, pp. 535–563, 1928.
[22] A. Rényi, "On measures of entropy and information," in *Proc. Fourth Berkeley Symp. Math. Stat. Prob., Vol.1: Contrib. to Theory Stat.*, 1961, pp. 547–561.
[23] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "ℓ-diversity: privacy beyond k-anonymity," in *Proc. 22nd Int. Conf. Data Eng.*, Atlanta, GA, 2006, pp. 24–24.
[24] C. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.
[25] L. Lovasz, "On the shannon capacity of a graph," *IEEE Trans. Inf. Theory*, vol. 25, no. 1, pp. 1–7, Jan. 1979.
[26] J. Korner and A. Orlitsky, "Zero-error information theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2207–2229, Oct. 1998.
[27] W. H. Cunningham, "Optimal attack and reinforcement of a network," *J. ACM*, vol. 32, no. 3, pp. 549–561, Jul. 1985.
[28] S. Fujishige, *Submodular functions and optimization*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 2005.
[29] A. V. Goldberg and R. E. Tarjan, "A new approach to the maximum-flow problem," *J. ACM*, vol. 35, no. 4, pp. 921–940, Oct. 1988.
[30] J. Hopcroft and R. Tarjan, "Algorithm 447: Efficient algorithms for graph manipulation," *Commun. ACM*, vol. 16, no. 6, pp. 372–378, Jun. 1973.
[31] E. Cheng and W. H. Cunningham, "A faster algorithm for computing the strength of a network," *Inf. Process. Lett.*, vol. 49, no. 4, pp. 209 – 212, Feb. 1994.
[32] V. Kolmogorov, "A faster algorithm for computing the principal sequence of partitions of a graph," *Algorithmica*, vol. 56, no. 4, pp. 394–412, 2010.
[33] S. Salamatian, A. Cohen, and M. Médard, "Maximum entropy functions: Approximate Gács-Körner for distributed compression," *arXiv preprint*, 2016. [Online]. Available: https://arxiv.org/abs/1604.03877
[34] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the information bottleneck to the privacy funnel," in *Proc. IEEE Inf. Theory Workshop*, Hobart, TAS, 2014, pp. 501–505.
[35] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," *arXiv preprint*, 2000. [Online]. Available: https://arxiv.org/abs/physics/0004057
[36] N. Slonim and N. Tishby, "Agglomerative information bottleneck," in *Proc.12th Int. Conf. Neural Inf. Process. Syst.*, Denver, CO, 1999, pp. 617–623.
[37] J. Liao, L. Sankar, O. Kosut, and F. P. Calmon, "Robustness of maximal α-leakage to side information," *arXiv preprint*, 2019. [Online]. Available: https://arxiv.org/abs/1901.07105
[38] I. Issa, S. Kamath, and A. B. Wagner, "An operational measure of information leakage," in *2016 Ann. Conf. Inf. Sci. Syst.*, Princeton, NJ, 2016, pp. 234–239.
[39] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *2013 IEEE 54th Ann. Symp. Foundations Comput. Sci.*, Berkeley, CA, 2013, pp. 429–438.
[40] A. D. Sarwate and L. Sankar, "A rate-disortion perspective on local differential privacy," in *Proc. 52th Ann. Allerton Conf. Comm. Control Comput.*, Monticello, IL, Sep. 2014, pp. 903–908.
[41] B. Rassouli, F. E. Rosas, and D. Gündüz, "Data disclosure under perfect sample privacy," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 2012–2025, Nov. 2020.
[42] H. Wang, M. Diaz, F. P. Calmon, and L. Sankar, "The utility cost of robust privacy guarantees," in *Proc. IEEE Int. Symp. Inf. Theory*, Vail, CO, 2018, pp. 706–710.
[43] X. Xu and S. Huang, "On the asymptotic sample complexity of HGR maximal correlation functions in semi-supervised learning," in *Proc. 57th Ann. Allerton Conf. Comm. Control Comput.*, Monticello, IL, 2019, pp. 879–886.
[44] S. Salamatian, A. Cohen, and M. Médard, "Efficient coding for multi-source networks using Gács-Körner common information," in *Proc. 2016 Int. Sym. Inf. Theory Its Appl.*, Monterey, CA, 2016, pp. 166–170.
[45] B. Rassouli and D. Gunduz, "On perfect privacy," in *Proc. IEEE Int. Symp. Inf. Theory*, Vail, CO, 2018, pp. 2551–2555.
[46] F. P. Calmon, A. Makhdoumi, and M. Médard, "Fundamental limits of perfect privacy," in *Proc. 2015 IEEE Int. Symp. Inform. Theory*, Hong Kong, 2015, pp. 1796–1800.
[47] C.-L. Hwang and A. S. M. Masud, *Multiple objective decision making, methods and applications: a state-of-the-art survey*. New York: Springer Science & Business Media, 2012, vol. 164.
[48] A. Asuncion and D. Newman, "UCI machine learning repository," 2007. [Online]. Available: https://archive.ics.uci.edu/ml/index.php
[49] L. Sweeney, "Achineving k-anonymity privacy protection using generalization and suppresion," *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.*, vol. 10, no. 05, pp. 571–588, 2002.
[50] N. Ding and P. Sadeghi, "A submodularity-based clustering algorithm for the information bottleneck and privacy funnel," in *Proc. IEEE Inf. Theory Workshop*, Visby, Sweden, 2019, pp. 1–5.
[51] F. Bach, "Convex analysis and optimization with submodular functions: a tutorial," *arXiv preprint arXiv:1010.4207*, 2010.
[52] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.
[53] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proc. IEEE*, vol. 86, no. 11, pp. 2210–2239, Nov. 1998.
[54] O. Shamir, S. Sabato, and N. Tishby, "Learning and generalization with the information bottleneck," *Theor. Comput. Sci.*, vol. 411, no. 29, pp. 2696 – 2711, 2010.
[55] A. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.



Ni Ding received the B.C.A. degree from the Shanghai Second Polytechnic University, China, the B.E. degree in Telecommunications (1st class honours) from the University of New South Wales, Australia, and the PhD degree from the Australian National University, Australia, in 2005, 2012 and 2017, respectively. She was a postdoctoral fellow at Data 61, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, from 2017 to 2019. She is now a Doreen Thomas Postdoctoral Fellow at the School of Computing and Information Systems, University of Melbourne. Her research interests generally include optimizations in information theory, wireless communications, signal processing and machine learning. She is currently interested in data privacy, discrete and combinatorial optimization problems raised in discrete event control in cross-layer adaptive modulation, source coding and game theory (in particular, the games with strong structures, e.g., supermodular and convex games).