Differential Privacy: on the trade-off between Utility and Information Leakage

Mário S. Alvim, Miguel E. Andrés, Konstantinos Chatzikokolakis, Pierpaolo Degano, Catuscia Palamidessi

Differential Privacy: on the trade-off between Utility and Information Leakage*

Mário S. Alvim[1], Miguel E. Andrés[1], Konstantinos Chatzikokolakis[1], Pierpaolo Degano[2], and Catuscia Palamidessi[1]

[1] INRIA and LIX, Ecole Polytechnique, France.
[2] Dipartimento di Informatica, Università di Pisa, Italy.

Abstract. Differential privacy is a notion of privacy that has become very popular in the database community. Roughly, the idea is that a randomized query mechanism provides sufficient privacy protection if the ratio between the probabilities that two adjacent datasets give the same answer is bound by e^{ϵ}. In the field of information flow there is a similar concern for controlling information leakage, i.e. limiting the possibility of inferring the secret information from the observables. In recent years, researchers have proposed to quantify the leakage in terms of Rényi min mutual information, a concept strictly related to the Bayes risk. In this paper, we show how to model the query system in terms of an information-theoretic channel, and we compare the notion of differential privacy with that of mutual information. We show that differential privacy implies a bound on the mutual information, but not vice-versa. Furthermore, we show that our bound is tight. Then, we consider the utility of the randomization mechanism, which represents how close the randomized answers are, in average, to the real ones. We show that the notion of differential privacy implies a bound on utility, also tight, and we propose a method that under certain conditions builds an optimal randomization mechanism, i.e. a mechanism which provides the best utility while guaranteeing ϵ-differential privacy.

1 Introduction

The area of statistical databases has been one of the first communities to consider the issues related to the protection of information. Already some decades ago, Dalenius [11] proposed a famous "ad omnia" privacy desideratum: nothing about an individual should be learnable from the database that cannot be learned without access to the database.

1.1 Differential privacy

Dalenius' property is too strong to be useful in practice: it has been shown by Dwork [12] that no useful database can provide it. In replacement Dwork

* This work has been partially supported by the project ANR-09-BLAN-0169-01 PANDA and by the INRIA DRI Equipe Associée PRINTEMPS.

has proposed the notion of *differential privacy*, which has had an extraordinary impact in the community. Intuitively, such notion is based on the idea that the presence or the absence of an individual in the database, or its particular value, should not change in a significant way the probability of obtaining a certain answer for a given query [12–15].

Dwork has also studied sufficient conditions for a randomized function $\mathcal{K}$ to implement a mechanism satisfying ϵ-differential privacy. It suffices to consider a Laplacian distribution with variance depending on ϵ, and mean equal to the correct answer [14]. This is a technique quite diffused in practice.

1.2 Quantitative information flow

The problem of preventing the leakage of secret information has been a pressing concern also in the area of software systems, and has motivated a very active line of research called *secure information flow*. Similarly to the case of privacy, also in this field, at the beginning, the goal was ambitious: to ensure *non-interference*, which means complete lack of leakage. But, as for Dalenius' notion of privacy, non-interference is too strong for being obtainable in practice, and the community has started exploring weaker notions. Some of the most popular approaches are the quantitative ones, based on information theory. See for instance [6, 7, 9, 18–20, 23].

The various approaches in literature differ, mainly, for the notion of entropy. These notions are related to the kind of attackers we want to model, and to how we measure their success (see [18] for an illuminating discussion of this relation). Shannon entropy [22], on which most of the approaches are based, is used to model an adversary which tries to find out the secret x by asking questions of the form "does x belong to set S?". Shannon entropy is precisely the average number of questions necessary to find out the exact value of x with an optimal strategy (i.e. an optimal choice of the S's). The other most popular notion of entropy (in this area) is Rényi's min entropy [21]. The corresponding notion of attack is a *single try* of the form "is x equal to v?". Rényi's min entropy is precisely the log of the probability of guessing the true value with the optimal strategy, which consists, of course, in selecting the v with the highest probability. Approaches based on this notion include [23] and [4].

In this paper, we focus on the approach based on the Rényi min entropy.

It is worth noting that, while the Rényi's min entropy of X, $H_\infty(X)$, represents the a priori probability of success (of the single-try attack), the Rényi's min conditional entropy of X given Y, $H_\infty(X \mid Y)$, represents the a posteriori probability of success[1]. This a posteriori probability is the converse of the Bayes risk [10], which has also been used as a measure of the leakage of secret information [3, 5].

[1] We should mention that Rényi did not define the conditional version of the min entropy, and that there have been various different proposals in literature for this notion. We use here the one proposed by Smith in [23].

1.3 Goal of the paper

The first goal of this paper is to explore the relation between differential privacy and quantitative information flow. We address the problem of characterizing the protection that differential privacy provides with respect to information leakage. Then, we consider the problem of the utility. This is different from information leakage in that it represents the relation between the reported answer and the true answer. While we want to avoid that the system leaks the information of the participants, we do not need the same protection towards the true answer in itself. It is therefore interesting to explore ways to improve the utility while preserving privacy. We attack this problem by considering the possible structure that the query induces on the true answers.

1.4 Contribution

The main contribution of this paper is the following

- We propose a model-theoretic framework to reason about both information leakage and utility.
- We prove that ϵ-differential privacy implies a bound on the information leakage. The bound is tight.
- We prove that ϵ-differential privacy implies a bound on the utility. We prove that, under certain conditions, the bound is tight.
- We identify a method that, under certain conditions, constructs the randomization mechanisms which maximizes utility while providing ϵ-differential privacy.

1.5 Plan of the paper

Next section introduces some necessary background notions. Section 3 proposes an information-theoretic view of the database query systems, and of its decomposition in terms of the query and of the randomization mechanisms. Section 4 shows that differential privacy implies a bound on the Rényi min mutual information, and that the bound is tight. Section 5 shows that differential privacy implies a bound on the utility, and that under certain conditions the bound is tight. Furthermore it shows how to construct and optimal randomization mechanism. Section 6 discusses related work, and Section 7 concludes.

The proofs of the results are in the appendix.

2 Background

2.1 Differential privacy

Roughly, the idea of differential privacy is that a randomized query mechanism provides sufficient privacy protection if the ratio between the probabilities of two different entries to originate a certain answer is bound by e^{ϵ}, for some given $\epsilon \geq 0$. Dwork's definition of differential privacy is the following:

Definition 1 ([14]). *A randomized function $\mathcal{K}$ satisfies* ϵ-differential privacy *if for all of data sets D' and D'' differing on at most one row, and all $S \subseteq Range(\mathcal{K})$,*

$$Pr[\mathcal{K}(D') \in S] \leq e^{\epsilon} \times Pr[\mathcal{K}(D'') \in S] \tag{1}$$

2.2 Information theory and interpretation in terms of attacks

In the following, X, Y denote two discrete random variables with carriers $\mathcal{X} = \{x_0, \ldots, x_{n-1}\}$, $\mathcal{Y} = \{y_0, \ldots, y_{m-1}\}$, and probability distributions $p_X(\cdot)$, $p_Y(\cdot)$, respectively. An information-theoretic channel is constituted by an input X, an output Y, and the matrix of conditional probabilities $p_{Y|X}(\cdot \mid \cdot)$, where $p_{Y|X}(y \mid x)$ represent the probability that Y is y given that X is x. We shall omit the subscripts on the probabilities when they are clear from the context.

Rényi min-entropy In [21], Rényi introduced an one-parameter family of entropy measures, intended as a generalization of Shannon entropy. The Rényi entropy of order α ($\alpha > 0$, $\alpha \neq 1$) of a random variable X is defined as $H_\alpha(X) = \frac{1}{1-\alpha} \log_2 \sum_{x \in \mathcal{X}} p(x)^\alpha$. We are particularly interested in the limit of H_α as α approaches ∞. This is called *min-entropy*. It can be proven that $H_\infty(X) \stackrel{\text{def}}{=} \lim_{\alpha \to \infty} H_\alpha(X) = -\log_2 \max_{x \in \mathcal{X}} p(x)$.

Rényi defined also the α-generalization of other information-theoretic notions, like the Kullback-Leibler divergence. However, he did not define the α-generalization of the conditional entropy, and there is no agreement on what it should be. For the case $\alpha = \infty$, we adopt here the definition of conditional entropy proposed by Smith in [23]:

$$H_\infty(X \mid Y) = -\log_2 \sum_{y \in \mathcal{Y}} p(y) \max_{x \in \mathcal{X}} p(x \mid y) \tag{2}$$

Analogously to the Shannon case, we can define the Rényi-mutual information I_∞ as $H_\infty(X) - H_\infty(X \mid Y)$, and the capacity C_∞ as $\max_{p_X(\cdot)} I_\infty(X;Y)$. It has been proven in [4] that C_∞ is obtained at the uniform distribution, and that it is equal to the sum of the maxima of each column in the channel matrix, i.e., $C_\infty = \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} p(y \mid x)$.

Interpretation in terms of attacks: Rényi min-entropy can be related to a model of adversary who is allowed to ask exactly one question, which must be of the form "is $X = x$?" (one-try attacks). More precisely, $H_\infty(X)$ represents the (logarithm of the inverse of the) probability of success for this kind of attacks and with the best strategy, which consists, of course, in choosing the x with the maximum probability.

As for $H_\infty(X \mid Y)$, it represents the inverse of the (expected value of the) probability that the same kind of adversary succeeds in guessing the value of X *a posteriori*, i.e. after observing the result of Y. The complement of this probability is also known as *probability of error* or *Bayes risk*. Since in general X and Y

are correlated, observing Y increases the probability of success. Indeed we can prove formally that $H_\infty(X \mid Y) \leq H_\infty(X)$, with equality if and only if X and Y are independent. $I_\infty(X;Y)$ corresponds to the *ratio* between the probabilities of success a priori and a posteriori, which is a natural notion of leakage. Note that $I_\infty(X;Y) \geq 0$, which seems desirable for a good notion of leakage.

3 A model of utility and privacy for statistical databases

In this section we present a model of statistical queries on databases, where noise is carefully added to protect privacy and, in general, the reported answer to a query does not need to correspond to the real one. In this model, the notion of information leakage can be used to measure the amount information that an attacker learns about the database. Moreover, the model allows us to quantify the utility of the query, that is, how much information about the real answer can be obtained from the reported one. This model will serve as the basis for exploring the relation between differential privacy and information flow.

We fix a finite set Ind of individuals that participate in the database and a finite set of possible values Val for each individual.[2] Let $u = |Ind|$ and $v = |Val|$. A database $D = \{d_0, \ldots, d_{u-1}\}$ is a u-tuple where each $d_i \in Val$ is the value of the corresponding individual. The set of all databases is $\mathcal{X} = Val^u$. Two databases D, D' are *adjacent*, written $D \sim D'$ iff they differ for the value of exactly one individual.

Let $\mathcal{K}$ be a randomized function and $\mathcal{Z} = Range(\mathcal{K})$. This function can be modelled by a channel $C_\mathcal{K}$ with input and output alphabets $\mathcal{X}, \mathcal{Z}$ respectively. This channel, displayed in Figure 1, can be specified as usual by a matrix of conditional probabilities $p_{Z|X}(\cdot|\cdot)$. We also denote by X, Z the random variables modelling the input and output of the channel. The definition of diffential privacy can be directly expressed as a property of the channel: $C_\mathcal{K}$ satisfies ϵ-differential privacy iff

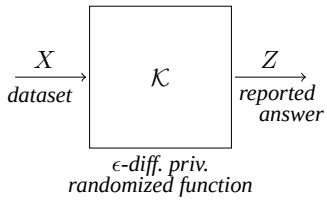


Fig. 1. A randomized function $\mathcal{K}$

$$p(z|x) \leq e^\epsilon p(z|x') \text{ for all } z \in \mathcal{Z}, x, x' \in \mathcal{X} \text{ with } x \sim x'$$

Intuitively, the *correlation* between X and Z measures how much information about the complete database the attacker can obtain by observing the reported answer. We will refer to this correlation as the *leakage* of the channel, denoted by $\mathcal{L}(X, Z)$. In Section 4 we discuss how this leakage can be quantified, using notions from information theory, and we study the behavior of the leakage for differentially private queries.

[2] The absence of an individual from the database, if allowed, can be represented by one of the values in Val. See the discussion at the end of this section.

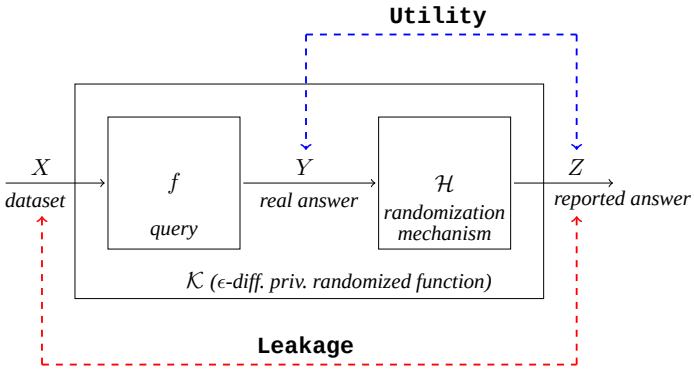


Fig. 2. Leakage and utility for oblivious mechanisms

We then introduce a random variable Y modelling the true answer to the query f, ranging over $\mathcal{Y} = Range(f)$. Now the correlation between Y and Z measures how much we can learn about the real answer from the reported one. We will refer to this correlation as the *utility* of the channel, denoted by $\mathcal{U}(Y, Z)$. In Section 5 we discuss in detail how utility can be quantified, and we investigate how to construct a private mechanism so that utility is maximized while preserving differential privacy.

In practice, the privacy mechanism that adds noise to the query output is often *oblivious*, meaning that the reported answer Z only depends on the real answer Y and not on the database X. In this case, the channel $C_{\mathcal{K}}$ can be decomposed into two parts: a channel C_f modelling the query f, and a channel $C_{\mathcal{H}}$ modelling the oblivious noise. The definition of utility in this case is simplified as it only depends on properties of the sub-channel $C_{\mathcal{H}}$. The leakage and the utility for a decomposed randomized function are displayed in Figure 2.

Leakage about an individual. As already discussed, $\mathcal{L}(X, Z)$ can be used to quantify the information that the attacker can learn about the whole database. However, protecting the whole database is not the main goal of differential privacy (indeed, some information will necessarily be revealed, otherwise the query would not be useful). Instead, differential privacy aims at protecting the value of an individual, even in the worst case where the values of all other individuals are known. To quantify this information leakage we can define smaller channels, where only the information of a specific individual varies. Let $D^- \in Val^{u-1}$ be a $(u-1)$-tuple with the values of all individuals but one. We create channel C_{D^-} whose input alphabet is the set of all databases in which the $u-1$ individuals have the same values as in D^-. Intuitively, the information leakage of this channel measures how much information about one individual the attacker can learn if the values of all others are known to be D^-. This leakage is studied in Section 4.1.

A note on the choice of values. The choice of the set Val depends on the assumptions about the attacker's knowledge. In particular, if the attacker does not know which individuals participate in the database, one of the values in Val (e.g. 0 or a special value *null*) could be interpreted as absence. As discussed in [14], a database D' adjacent to D can be though of either being a superset of D with one extra row, or the same as D in all rows except from one. Our definition of $\sim$ with the posibility of *null* values covers both cases.

However, an important observation should be made about the choice of Val. Most often we are interested in protecting the *actual value* of an individual, not just its participation in the database. In this case, the definition of differential privacy (as well as the channels we are constructing) should include databases with all possible values for each individual, not just the "real" ones. In other words, to prevent the attacker from finding out the individual's value, the probability $p(z|x)$, where x contains the individual's true value, should be close to $p(z|x')$ where x' contains a hypothetical value for this individual.

This might seem unnecessary at first sight, since differential privacy is often though as protecting an individual's participation in a database. However, hiding an individual's participation does not imply hiding his value. Consider the following example: we aim at learning the average salary of employees in a small company, and it happens that all of them have exactly the same salary s. We allow anyone to participate or not, while offering ϵ-differential privacy. If we only consider s as the value in all possible databases, then the query is always constant, so answering it any number of times without any noise should satisfy differential privacy for any $\epsilon > 0$. Since all reported answers are s, the attacker can deduce that the salary of all employees (including those not participating in the query) is s. In fact, the attacker cannot find out who participated, despite the value of all individuals is revealed.

In other cases, of course, the values are public and we are only interested in hiding the participation (e.g. average height of people with cancer). Thus, Val should be properly selected according to the application. If participation is known and we only wish to hide the values, Val should contain all possible values. If the values are known and participation should be hidden, Val could be $\{0, 1\}$ denoting absence or presence respectively. Finally, if both the value and the participation should be protected, Val can contain all values plus *null*.

4 Leakage

As discussed in the previous section, the correlation $\mathcal{L}(X, Z)$ between X and Z measures the information that the attacker can learn about the database. In this section, we consider Rényi min-entropy as a measure of this leakage, that is $\mathcal{L}(X, Z) = I_\infty(X; Z)$. We then investigate bounds on information leakage imposed by differential privacy.

Our first result shows that the leakage of a randomized function $\mathcal{K}$ is bounded by a quantity depending on ϵ, the numbers u, v of individuals and values respectively. We assume that $v \geq 2$.

Theorem 1. *If $\mathcal{K}$ provides* ϵ-differential privacy *then the leakage associated to $\mathcal{K}$ is bounded from above as follows:*

$$I_\infty(X;Z) \leq u \log_2 \frac{v\, e^\epsilon}{(v-1+e^\epsilon)}$$

Note that this bound $B(u,v,\epsilon) = u \log_2 \frac{v\, e^\epsilon}{(v-1+e^\epsilon)}$ is a continuous function in ϵ, has value 0 when $\epsilon = 0$, and converges to $u \log_2 v$ as ϵ approaches infinity. Figure 3 shows the growth of B with respect to ϵ, for various fixed values of u and v.

The following result shows that the bound $B(u,v,\epsilon)$ is *tight*.

Proposition 1. *For every u, v, and ϵ there exists a randomized function $\mathcal{K}$ which provides ϵ-differential privacy and whose leakage, for the uniform input distribution, is $I_\infty(X;Z) = B(u,v,\epsilon)$.*

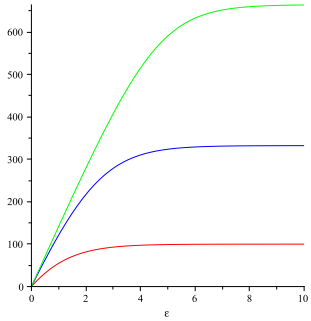
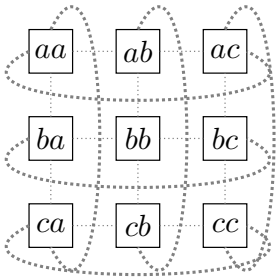


Fig. 3. Graphs of $B(u,v,\epsilon)$ for $u = 100$ and $v = 2$ (lowest line), $v = 10$ (intermediate line), and $v = 100$ (highest line), respectively.

Example 1. Assume that we are interested in the eyes color of a certain population $Ind = \{Alice, Bob\}$. Let $Val = \{a, b, c\}$ where a stands for *absent*, b for *blue*, and c for *chatain*. We can represent each dataset with a tuple $d_1 d_0$, where $d_0 \in V$ represent the eyes color of *Alice* (cases $d_0 = b$ and $d_0 = c$), or that *Alice* is not in the dataset (case $d_0 = a$). d_1 provides the same kind of information for *Bob*. Note that $v = 3$. Fig 4(a) represents the set $\mathcal{X}$ of all possible datasets and its adjacency relation. Fig 4(b) represents the matrix with input $\mathcal{X}$ which provides ϵ-differential privacy and has the highest information leakage. In the representation of the matrix, the generic u stands for $\frac{a}{e^{\epsilon\, u}}$, where a is the highest value in the matrix, i.e. $a = \frac{v\, e^\epsilon}{(v-1+e^\epsilon)} = \frac{3\, e^\epsilon}{(2+e^\epsilon)}$.

We know from the literature [4, 23] that the I_∞ of a given matrix has its maximum in correspondence of the uniform input distribution. So, it is natural to ask whether we could find a tighter bound on the leakage when we assume a fixed (non-uniform) input distribution. Furthermore, the construction of the matrix for Proposition 1 gives a square matrix of dimension $v^u \times v^u$. Often, however, the range of $\mathcal{K}$ is fixed, as it is usually related to the possible answers to the query f. Hence it is natural to consider the scenario in which we are given a number $r < v^u$, and want to consider only those $\mathcal{K}$'s whose range has cardinality at most r. Could we, in this restricted setting, find a better bound than the one given by Theorem 1? The following proposition answers these questions.

In order to state the proposition as simply as possible, it is convenient to index the input probabilities according to a decreasing order: $p_0 = p(x_0) = \max_{\mathcal{X}} p(x) \geq p_1 = p(x_1) = \max_{\mathcal{X}\setminus\{x_0\}} p(x) \geq p_2 = p(x_2) = \max_{\mathcal{X}\setminus\{x_0,x_1\}} p(x)$, etc.



(a) The datasets and their adjacency relation

	aa	ab	ac	ba	ca	bb	bc	cb	cc
aa	0	1	1	1	1	2	2	2	2
ab	1	0	1	2	2	1	2	1	2
ac	1	1	0	2	2	2	1	2	1
ba	1	2	2	0	1	1	2	1	2
ca	1	2	2	1	0	2	2	1	1
bb	2	1	2	1	2	0	1	1	2
bc	2	2	1	1	2	1	0	2	1
cb	2	1	2	2	1	1	2	0	1
cc	2	2	1	2	1	2	1	1	0

(b) The representation of the matrix

Fig. 4. Universe and highest-leakage matrix giving ϵ-differential privacy for Example 1.

Proposition 2. *Let $\mathcal{K}$ be a randomized function and let $r = |Range(\mathcal{K})|$. If $\mathcal{K}$ provides* ϵ-differential privacy *then the leakage associated to $\mathcal{K}$ is bounded from above as follows:*

$$I_\infty(X;Z) \leq -\log_2 p_0 + \max_{1 \leq k \leq \lfloor \log_v r \rfloor} k \log_2 \left(\frac{e^\epsilon}{(v-1+e^\epsilon)} \sum_{h=0}^{s-1} p_h \right)$$

where $s = \min\{r, v^{k+1}\}$.

Note that this bound can be much smaller than the one provided by Theorem 1. For instance, if $r = s$, and the input probability is uniform, this bound can be at most $-\log_2 \frac{1}{v^u} + \log_2 \left(\frac{e^\epsilon}{(v-1+e^\epsilon)} \sum_{h=0}^{v-1} \frac{1}{v^u} \right) = \log_2 \frac{v\, e^\epsilon}{(v-1+e^\epsilon)}$, i.e. u times smaller than $B(u, v, \epsilon)$. Let us clarify that there is no contradiction with the fact that the bound $B(u, v, \epsilon)$ is strict: in fact it is strict when we are free to choose the range, but here we fix the dimension of the range.

Also a non-uniform input distribution can help lowering the bound. This is obvious, because the bound provided by Proposition 2 is always smaller than $-\log_2 p_0$ (independently from u and v). This is because $v \geq 2$ hence the argument of the second logarithm is smaller than 1.

4.1 Measuring the leakage about an individual

As discussed in Section 3, the main goal of differential privacy is not to protect information about the complete database, but about each individual. To capture the information leakage about an individual, we start from a tuple $D^- \in Val^{u-1}$ containing the exact values of all other individuals. Then we create a channel whose input X_{D^-} ranges over all databases where the values of the $u-1$ are exactly those of D^- and only the value of one individual varies. Intuitively, $I_\infty(X_{D^-}; Y)$ measures the leakage about the individual's value where all other values are known to be D^-. As all these databases are adjacent, differential privacy provides a stronger bound for this leakage.

Theorem 2. *If $\mathcal{K}$ provides* ϵ-differential privacy *then for all $D^- \in Val^{u-1}$ the leakage about an individual is bounded from above as follows:*

$$I_\infty(X_{D^-}; Z) \leq \log_2 e^\epsilon$$

Note that this bound is stronger than the one of Theorem 1. In particular, it depends only on ϵ and not on u, v.

5 Utility

As discussed in Section 3, the utility of a randomized function $\mathcal{K}$ is the correlation between the real answers Y for a query and the reported answers Z. In this section we analyze the utility $\mathcal{U}(Y, Z)$ using the classic notion of *utility functions* (see for instance [2]).

For our analysis we assume that an oblivious privacy mechanism is employed. As discussed in Section 3, in this case the system can be decomposed into two channels, and the utility becomes a property of the channel $C_\mathcal{H}$ which maps the real answer $y \in \mathcal{Y}$ into a reported answer $z \in \mathcal{Z}$ according to probability distributions $p_{Z|Y}(\cdot|\cdot)$. However, the user does not necessarily take z as her guess for the real answer, since she can use some Bayesian post-processing to maximize the probability of a right guess. Thus for each reported answer z the user can remap her guess to a value $y' \in \mathcal{Y}$ according to some strategy that maximizes her gain. For each pair of (y, y') there is an associated value given by a gain (or utility) function $g(y, u)$ that represents how much the user gains by guessing y' when the real answer is y.

It is natural to define the global utility of the mechanism $\mathcal{H}$ as the expected gain:

$$\mathcal{U}(Y, Z) = \sum_y p(y) \sum_{y'} p(y'|y) g(y, y') \tag{3}$$

where $p(y)$ is the prior probability of real answer y, $p(y'|y)$ is the probability of user guessing y' when the real answer is y, and $g(y, y')$ is the gain function of the pair (y, y').

Assuming that the user uses a remapping function $\rho(z) : \mathcal{Z} \to \mathcal{Y}$, we can derive the following characterization of the utility. We will use δ_x to represent

the probability distribution which has value 1 on x and 0 elsewhere.

$$\begin{aligned}
\mathcal{U}(Y,Z) &= \sum_y p(y) \sum_{y'} p(y'|y) g(y,y') && \text{(by (3))} \\
&= \sum_y p(y) \sum_{y'} \left(\sum_z p(z|y) p(y'|z) \right) g(y,y') && \text{(by remap)} \\
&= \sum_y p(y) \sum_{y'} \left(\sum_z p(z|y) \delta_{\rho(z)}(y') \right) g(y,y') && \text{(the remap is a function)} \\
&= \sum_y p(y) \sum_z p(z|y) \sum_{y'} \delta_{\rho(z)}(y') g(y,y') \\
&= \sum_{y,z} p(y,z) \sum_{y'} \delta_{\rho(z)}(y') g(y,y') \\
&= \sum_{y,z} p(y,z) g(y,\rho(z))
\end{aligned}$$

A very common utility function is the *binary gain function* defined as $g_{\text{bin}}(y,y') = 1$ if $y = y'$ and $g_{\text{bin}}(y,y') = 0$ if $y \neq y'$. In the rest of this section we will focus on the binary case. The use of binary utility functions in the context of differential privacy was also investigated in [16][3].

By substituting g with g_{bin} in the above formula we obtain:

$$\mathcal{U}(Y,Z) = \sum_{y,z} p(y,z) \delta_y(\rho(z)) \tag{4}$$

which tells us that the expected utility is the greatest when $\rho(z) = y$ is chosen to maximize $p(y,z)$. Assuming that the user chooses such a maximizing remapping, we have:

$$\mathcal{U}(Y,Z) = \sum_{y,z} \max_y p(y,z) \tag{5}$$

This corresponds to the converse of the Bayes risk, and it is closely related to the Rény min conditional entropy and to the Rény min mutual information:

$$H_\infty(Y|Z) = -\log_2 \mathcal{U}(Y,Z) \qquad I_\infty(Y;Z) = H_\infty(X) + \log_2 \mathcal{U}(Y,Z)$$

5.1 A bound on the utility

In this section we show that the fact that $\mathcal{K}$ provides ϵ-differential privacy induces a bound on the utility. We start by extending the adjacency relation $\sim$ from the datasets to the answers $\mathcal{Y}$. Intuitively, the function f associated to the query determines a partition on the set of all databases ($\mathcal{X}$, i.e. Val^u), and we say that two classes are adjacent if they contain an adjacent pair. More formally:

[3] The authors of [16] used the dual notion of *loss functions* instead than gain functions, but the final result is equivalent.

Definition 2. *Given $y, y' \in \mathcal{Y}$, with $y \neq y'$, we say that y and y' are adjacent (notation $y \sim y'$), iff there exist $D, D' \in Val^u$ with $D \sim D'$ such that $y = f(D)$ and $y' = f(D')$.*

Since $\sim$ is symmetric on databases, it is also symmetric on $\mathcal{Y}$, therefore $(\mathcal{Y}, \sim)$ forms an undirected graph. We define the distance *dist* between two elements $y, y' \in \mathcal{Y}$ as the length of the minimum path from y to y'. For a given natural number d, we use $Border(y, d)$ to denote the set of elements at distance d from y, i.e.

$$Border(y, d) = \{y' \mid dist(y, y') = d\}$$

We recall that a graph automorphism σ is an isomorphism of the graph into itself. In a (finite) graph, an orbit of σ is a set of the form $\{v, \sigma(v), \sigma^2, \ldots, \sigma^{n-1}(v)\}$ where v is an arbitrary vertex and $\sigma^n(v) = v$.

We are now ready to give a bound on the utility:

Theorem 3. *Let $\mathcal{H}$ be a randomization mechanism for the randomized function $\mathcal{K}$ and the query f, and assume that $\mathcal{K}$ provides ϵ-differential privacy. Assume that $(\mathcal{Y}, \sim)$ admits a graph automorphism with only one orbit. Furthermore, assume that there exists a natural number c and an element $y \in \mathcal{Y}$ such that, for every d, either $|Border(y, d)| = 0$ or $|Border(y, d)| \geq c$. Then*

$$\mathcal{U}(X, Y) \leq \frac{(e^\epsilon)^n (1 - e^\epsilon)}{(e^\epsilon)^n (1 - e^\epsilon) + c\,(1 - (e^\epsilon)^n)}$$

where n is the maximum distance from y in $\mathcal{Y}$.

The bound provided by the above theorem is strict, in the sense that, when $|Border(y, d)|$ is exactly c for every d, then we can construct a $\mathcal{H}$ which has precisely that utility and it still provides ϵ-differential privacy. This randomization mechanism is therefore optimal, in the sense that it optimizes the utility for the given ϵ. This is the main topic of the next section.

5.2 Constructing an optimal randomization mechanism

Given a query f, and a differential privacy requirement ϵ, it is important to design the randomization mechanism $\mathcal{H}$ in such a way that (together with f) provides ϵ-differential privacy, but without making useless sacrifices on the utility. We shows a method to construct the optimal $\mathcal{H}$, at least in some particular cases.

Let f be a query and ϵ a differential privacy requirement. Assume that $(\mathcal{Y}, \sim)$ admits a graph automorphism with only one orbit. Assume that, for every $y \in Y$ and every natural number d, either $|Border(y, d)| = 0$ or $|Border(y, d)| = c$. Then we can construct an optimal randomization mechanism $\mathcal{H}$ in the following way. Let $\mathcal{Z} = \mathcal{Y}$ and $a = \frac{(e^\epsilon)^n (1 - e^\epsilon)}{(e^\epsilon)^n (1 - e^\epsilon) + c\,(1 - (e^\epsilon)^n)}$ where n is the diameter of the graph, namely the maximal distance between any two nodes. We define the matrix of conditional probabilities associated to $\mathcal{H}$ as follows: For every column z, define

$$p_{Z|Y}(z|y) = \frac{a}{(e^\epsilon)^d} \qquad \text{where } d = dist(y, z) \tag{6}$$

Theorem 4. *The definition in (6) determines a legal channel matrix for $\mathcal{H}$, i.e., for each y, $p_{Z|Y}(\cdot|y)$ is a probability distribution. Furthermore, it meets the differential privacy requirement, in the sense that the the composition $\mathcal{K}$ of f and $\mathcal{H}$ provides ϵ-differential privacy. Finally, $\mathcal{H}$ is optimal in the sense that it maximizes utility when the input distribution (i.e. the distribution of Y) is uniform.*

The conditions for the construction of the optimal matrix are strong, but there are some interesting scenarios in which they are satisfied. Depending on the degree of connectivity c, we can have $|\mathcal{Y}| - 2$ different cases (note that the case of $c = 1$ is not possible because the datasets are fully connected via their adjacency relation), whose extreme are:

- $(\mathcal{Y}, \sim)$ is a *ring*, i.e. every element has exactly two adjacent elements. This is similar to the case of the counting queries considered in [16], with the difference that our "counting" is in arithmetic modulo $|\mathcal{Y}|$.
- $(\mathcal{Y}, \sim)$ is a *clique*, i.e. every element has exactly $|\mathcal{Y}| - 1$ adjacent elements.

The optimal matrices generated by our algorithm above can be very different, depending on the value of c. Next examples illustrate two queries that give rise to the clique and to the ring structures, and show the corresponding matrices.

Example 2. Consider a database with electoral information where each row corresponds to a voter and contains (for simplicity) only three fields:

- ID: a unique (anonymized) identifier assigned to each voter;
- CITY: the name of the city on which the user voted;
- CANDIDATE: the name of the candidate the user voted for.

We will analyze two different queries for this database. First, consider the query *"What is the city with the greatest number of votes for a given candidate cand?"*. For such a query the binary function is a natural choice for the gain function: only the right city gives some gain, and any wrong answer is just as bad/good as any other. It is easy to see that every pairs of answers are neighbors, i.e. the graph structure of the answers is a clique.

Let us consider the scenario where CITY={A,B,C,D,E,F} and assume that there is a unique answer for the query, i.e., there are no two cities with exactly the same number of individuals voting for candidate *cand*. Table 1 shows two alternative mechanisms providing ϵ-differential privacy (with $\epsilon = \log 2$). The first one, M_1, is based on the truncated geometric mechanism method used in [16] for counting queries (here extended to the case where every paris of answers are neighbors). The second mechanism, M_2, is the one we propose in this paper.

Taking as input distribution the uniform distribution, it is easy to see that $\mathcal{U}(M_1) = 0.2243 < 0.2857 = \mathcal{U}(M_2)$. The gap becomes larger if we take an input distribution with lower values in the first and last row. For instance, for $p(A) = p(F) = 1/10$ and $p(B) = p(C) = p(D) = P(E) = 1/5$, we have $\mathcal{U}(M_1) = 0.1622 < 0.2857 = \mathcal{U}(M_2)$. This is not too surprising: the Laplacian method and

(a) M_1: truncated geometric mechanism

In/Out	A	B	C	D	E	F
A	0.535	0.060	0.052	0.046	0.040	0.267
B	0.465	0.069	0.060	0.053	0.046	0.307
C	0.405	0.060	0.069	0.060	0.053	0.353
D	0.353	0.053	0.060	0.069	0.060	0.405
E	0.307	0.046	0.053	0.060	0.069	0.465
F	0.268	0.040	0.046	0.052	0.060	0.534

(b) M_2: our mechanism

In/Out	A	B	C	D	E	F
A	2/7	1/7	1/7	1/7	1/7	1/7
B	1/7	2/7	1/7	1/7	1/7	1/7
C	1/7	1/7	2/7	1/7	1/7	1/7
D	1/7	1/7	1/7	2/7	1/7	1/7
E	1/7	1/7	1/7	1/7	2/7	1/7
F	1/7	1/7	1/7	1/7	1/7	2/7

Table 1. Mechanisms for the city with higher number of votes for candidate *cand*

the geometric mechanism work very well when the domain of answers is provided with a metric and the utility function is not binary[4]. It also works well when $(\mathcal{Y}, \sim)$ has low connectivity, in particular in the cases of a ring and of a line. But in this example, we are not in these cases, because we are considering *binary gain functions* and *high connectivity*.

Example 3. Let us consider the same database as the previous example, but now assume a counting query of the form *"What is the number of votes for candidate cand?"*. It is easy to see that each answer has at most two neighbors. More precisely, the graph structure on the answers is a line. For illustration purposes, let us assume that only 5 individuals have participated on the election. Table 2 shows two alternative mechanisms providing ϵ-differential privacy ($\epsilon = \log 2$): the truncated geometric mechanism method M_1 proposed in [16] and the mechanism we propose M_2, where $c = 2$ and $n = 3$. (Note that we can use our mechanism also when $|Border(y, d)| \leq c$, since it still satisfies differential privacy. However in this case it is not guaranteed to be optimal.)

In addition, suppose that the user querying the database has as prior information that the extreme cases (where all five participants voted for *cand* and none of them voted for *cand*, respectively) have low probability: $p(A) = p(F) = 1/100$, $p(B) = p(D) = 24/100$ and $p(C) = P(D) = 25/100$. Simple calculations show that $\mathcal{U}(M_1) = 0.34 < 0.3636 = \mathcal{U}(M_2)$.

On the other hand, in case of uniform prior distribution the utility of M_1 is higher than the utility of M_2, in fact the first is $4/9$ and the second is $4/11$. This does not contradict our theorem, because our matrix is guaranteed to be optimal only in the case of a ring structure, not a line as we have in this example. If the structure were a ring, i.e. if the last row were adjacent to the first one, then M_1 would not provide ϵ-differential privacy.

[4] In the metric case the gain function can take into account the proximity of the reported answer to the real one, the idea being that a close answer, even if wrong, is better than a distant one.

(a) M_1: truncated $\frac{1}{2}$-geom. mechanism

In/Out	0	1	2	3	4	5
0	2/3	1/6	1/12	1/24	1/48	1/48
1	1/3	1/3	1/6	1/12	1/24	1/24
2	1/6	1/6	1/3	1/6	1/12	1/12
3	1/12	1/12	1/6	1/3	1/6	1/6
4	1/24	1/24	1/12	1/6	1/3	1/3
5	1/48	1/48	1/24	1/12	1/6	2/3

(b) M_2: our mechanism

In/Out	0	1	2	3	4	5
0	4/11	2/11	1/11	1/22	1/11	2/11
1	2/11	4/11	2/11	1/11	1/22	1/11
2	1/11	2/11	4/11	2/11	1/11	1/22
3	1/22	1/11	2/11	4/11	2/11	1/11
4	1/11	1/22	1/11	2/11	4/11	2/11
5	2/11	1/11	1/22	1/11	2/11	4/11

Table 2. Mechanisms for the counting query (5 voters)

6 Related work

Barthe and Köpf have investigated the connection between differential privacy and the Rényi min-entropy leakage in [1]. They also propose a characterization of differential privacy mechanisms in terms of information-theoretic channels, and they use an elegant representation of the inputs (which in their setting corresponds to the subsets of a given database) in terms of binary sequences. Then they proceed to show that the leakage of such channels depends on the size of the channel's input and provide a bound for the leakage. Our paper differs from theirs in the following aspects: (a) We provide a smaller bound for the information leakage, and we show that our bound is tight, in the sense that there exists a differentially private channel whose leakage achieves our bound. (b) Their approach captures only the particular (yet very interesting) case in which the focus of differential privacy is on hiding participation of individuals in a database. We consider both the absence/presence of the individuals, and, for those who are present, also their values (see discussion in Section 3). (c) They consider only the case of uniform prior distribution, while we provide a (smaller) bound for non-uniform prior. (d) In our model we also consider the case in which the randomized function is decomposed into a query and a randomization mechanism on the answers, so that we are able to separate the leakage and the utility, and study how to improve the utility without affecting (the bound on) the leakage.

Clarkson and Schneider also considered differential privacy as a case study of their proposal for quantification of integrity [8]. There, the authors analyze database privacy conditions from the literature (such as differential privacy, k-anonymity, and l-diversity) using their framework for utility quantification. In particular, they study the relationship between differential privacy and a notion of leakage (which is different from ours - in particular their definition is based on Shannon entropy) and they provide a tight bound on leakage.

Heusser and Malacaria [17] were among the first to explore the application of information-theoretic concepts to databases queries. They proposed to model database queries as programs, which allows for statical analysis of the information leaked by the query. However [17] did not attempt to relate information leakage to differential privacy.

In [16] the authors aim at obtaining optimal-utility randomization mechanisms while preserving differential privacy. The authors propose adding noise to the output of the query according to the geometric mechanism. Their framework is very interesting in the sense it provides a general definition of utility for a mechanism M that captures any possible side information and preference (defined as a loss function) the users of M may have. They prove that the geometric mechanism is optimal in the particular case of counting queries. Our results in Section 5 do not restrict to counting queries, but on the other hand we only consider the case of binary loss function.

7 Conclusion and future work

An important question in statistical databases is how to deal with the trade-off between the privacy offered to the individuals participating in the database and the offered utility. In this work we propose a model integrating the two notions of privacy and utility in the scenario where differential-privacy is applied. We provide a strict bound on the information leakage of a randomized function satisfying ϵ-differential privacy and, in addition, we study the utility of oblivious differential privacy mechanisms. We provide a way to optimize utility subject to differential privacy, in the scenario where a binary gain function is used to measure the utility of the answer to a query.

As future work, we want to generalize the bounds for more generic gain functions, possibly by using the Kantorovich metric to compare the a priori and a posteriori probability distributions on secrets.

References

1. Gilles Barthe and Boris Köpf. Information-theoretic bounds for differentially private mechanisms. In *Proc. of CSF*, 2011. To appear.
2. Jose M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. J. Wiley & Sons, Inc., 1994.
3. Christelle Braun, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Compositional methods for information-hiding. In *Proc. of FOSSACS*, volume 4962 of *LNCS*, pages 443–457. Springer, 2008.
4. Christelle Braun, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Quantitative notions of leakage for one-try attacks. In *Proc. of MFPS*, volume 249 of *ENTCS*, pages 75–91. Elsevier, 2009.
5. Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Prakash Panangaden. On the Bayes risk in information-hiding protocols. *J. of Comp. Security*, 16(5):531–571, 2008.
6. David Clark, Sebastian Hunt, and Pasquale Malacaria. Quantitative analysis of the leakage of confidential data. In *Proc. of QAPL*, volume 59 (3) of *Electr. Notes Theor. Comput. Sci*, pages 238–251. Elsevier, 2001.
7. David Clark, Sebastian Hunt, and Pasquale Malacaria. Quantitative information flow, relations and polymorphic types. *J. of Logic and Computation*, 18(2):181–199, 2005.

8. M. R. Clarkson and F. B. Schneider. Quantification of integrity, 2011. Tech. Rep.. `http://hdl.handle.net/1813/22012`.
9. Michael R. Clarkson, Andrew C. Myers, and Fred B. Schneider. Belief in information flow. *J. of Comp. Security*, 17(5):655–701, 2009.
10. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. J. Wiley & Sons, Inc., second edition, 2006.
11. Tore Dalenius. Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 15:429 — 444, 1977.
12. Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming, 33rd Int. Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proc., Part II*, volume 4052 of *LNCS*, pages 1–12. Springer, 2006.
13. Cynthia Dwork. Differential privacy in new settings. In *Proc. of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 174–183. SIAM, 2010.
14. Cynthia Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–96, 2011.
15. Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proc. of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 371–380. ACM, 2009.
16. Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proc. of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 351–360. ACM, 2009.
17. Jonathan Heusser and Pasquale Malacaria. Applied quantitative information flow and statistical databases. In *Formal Aspects in Security and Trust*, pages 96–110, 2009.
18. Boris Köpf and David A. Basin. An information-theoretic model for adaptive side-channel attacks. In *Proc. of CCS*, pages 286–296. ACM, 2007.
19. Pasquale Malacaria. Assessing security threats of looping constructs. In *Proc. of POPL*, pages 225–235. ACM, 2007.
20. Pasquale Malacaria and Han Chen. Lagrange multipliers and maximum information leakage in different observational models. In *Proc. of PLAS*, pages 135–146. ACM, 2008.
21. Alfréd Rényi. On Measures of Entropy and Information. In *Proc. of the 4th Berkeley Symposium on Mathematics, Statistics, and Probability*, pages 547–561, 1961.
22. Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 625–56, 1948.
23. Geoffrey Smith. On the foundations of quantitative information flow. In *Proc. of FOSSACS*, volume 5504 of *LNCS*, pages 288–302. Springer, 2009.

Appendix

In the following we assume that A and B are random variables with carriers $\mathcal{A}$ and $\mathcal{B}$, respectively. If M is a channel matrix of conditional probabilities $p_{B|A}(\cdot|\cdot)$, then we denote by $B(M, A)$ the random variable determined as output by channel M with input A. The conditional min-entropy $H_\infty(A, B(M, A))$ is denoted by $H_\infty^M(A)$. Similarly, $I_\infty(A, B(M, A))$ is denoted by $I_\infty^M(A)$. Also, for a matrix M, we denote by $\max_j^M$ the maximum element of column j over all rows i, i.e. $\max_j^M = \max_i M_{ij}$.

In the following lemmata we assume that A has a uniform distribution.

Lemma 1. *Given a channel matrix M with dimensions $n \times m$ ($m \geq n$) satisfying ϵ-differential privacy for some $\epsilon \geq 0$, we can construct a channel matrix M' with the same dimensions such that:*

1. *M' respects ϵ-differential privacy;*
2. *for every row i in M' there exists exactly one column j such that $M_{ij} = \max_j^M \neq 0$;*
3. *$I_\infty^{M'}(A) \geq I_\infty^M(A)$.*

Proof. For each row i of M, proceed as follows. If for row i there is exactly one j such that $M_{ij} = \max_j^M \neq 0$, just copy row i into as it is into the new matrix M'. If there are several $j_{k_1}, j_{k_2}, \ldots j_{k_q}$ such that $M_{j_{k_h}} = \max_{j_{k_h}}^M \neq 0$, then "collapse" columns $j_{k_1}, j_{k_2}, \ldots j_{k_q}$ as follows:

$$M'_{ij_{k_1}} = \sum_{h=1}^{q} M_{ij_{k_h}} \tag{7}$$

$$M'_{ij_{k_h}} = 0 \qquad \text{for } h \in \{k_2, k_3, \ldots, k_q\} \tag{8}$$

Note that if M respects ϵ-differential privacy, the sum the contents of columns j' and j'' respect $\frac{M_{i'j'}+M_{i'j''}}{M_{i''j'}+M_{i''j''}} \leq \frac{e^\epsilon M_{i''j'}+e^\epsilon M_{i''j''}}{M_{i''j'}+M_{i''j''}} \leq e^\epsilon$. As for the columns that were zero-ed, it is trivial that they respect a ϵ-differential privacy.

Moreover, it is easy to see that the value of $H_\infty^{M'}(A)$ cannot be smaller than the value of $H_\infty^M(A)$, since by construction the sum of the maxima of each column before and after the transformation is the same, i.e., $H_\infty^M(A) = \sum_j \max_j^M = \sum_j \max_j^{M'} = H_\infty^{M'}(A)$. As $H_\infty(A)$ is constant (A has the uniform distribution), it follows that $I_\infty^{M'}(A) = I_\infty^M(A)$.

□

Note that, if for every row i in M' there exists exactly one column j such that $M_{ij} = \max_j^M \neq 0$, then we can rearrange the columns of M' so that all the maxima are in the diagonal, i.e. $M'_{i,i} = \max_i^{M'}$ for all $i \in \mathcal{X}$. This is just a matter of representation, and it will be used in the next lemma to simplify the indexing.

Lemma 2. *Let M be a channel with input and output alphabets $\mathcal{X} = \mathcal{Z} = Val^u$ and a uniform input distribution. We fix the adjacency relation $\sim$ to the one defined in Section 3. Assume that the maximum value of each column is on the diagonal, that is $M_{i,i} = \max_i^M$ for all $i \in \mathcal{X}$. If M satisfies ϵ-differential privacy then we can construct a new channel matrix M' such that:*

1. *M' respects ϵ-differential privacy;*
2. *$M'_{i,i} = M'_{h,h}$ for all $i, h \in \mathcal{X}$ i.e. all the elements of the diagonal are equal;*
3. *$M'_{i,i} = \max_i^{M'}$ for all $i \in \mathcal{X}$;*

4. $I_\infty^M(X) = I_\infty^{M'}(X)$.

Proof. Let $k, l \in Val^u$. We denote by $d(k,l)$ the number of elements in which k, l differ, which is the length of the minimum $\sim$-path connecting k and l. Since $\mathcal{X} = \mathcal{Z} = Val^u$ we will use $d(\cdot,\cdot)$ also between rows and columns. We also define $\mathcal{Z}_{h,d} = \{k \in \mathcal{Z} | d(h,k) = d\}$.

Let $n = |\mathcal{X}| = v^u$. The matrix M' is given by

$$M'_{hk} = \frac{1}{n|\mathcal{Z}_{h,d(h,k)}|} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{Z}_{i,d(h,k)}} M_{ij}$$

We first show that this is a well defined channel matrix, namely $\sum_{k \in \mathcal{Z}} M'_{hk} = 1$ for all $h \in \mathcal{X}$. We have

$$\begin{aligned}\sum_{k \in \mathcal{Z}} M'_{hk} &= \sum_{k \in \mathcal{Z}} \frac{1}{n|\mathcal{Z}_{h,d(h,k)}|} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{Z}_{i,d(h,k)}} M_{ij} \\ &= \frac{1}{n} \sum_{i \in \mathcal{X}} \sum_{k \in \mathcal{Z}} \frac{1}{|\mathcal{Z}_{h,d(h,k)}|} \sum_{j \in \mathcal{Z}_{i,d(h,k)}} M_{ij}\end{aligned}$$

Let $\Delta = \{0, \ldots, u\}$. Note that $\mathcal{Z} = \bigcup_{d \in \Delta} \mathcal{Z}_{h,d}$, and these sets are disjoint, so the summation over $k \in \mathcal{Z}$ can be split as follows

$$\begin{aligned}&= \frac{1}{n} \sum_{i \in \mathcal{X}} \sum_{d \in \Delta} \sum_{k \in \mathcal{Z}_{h,d}} \frac{1}{|\mathcal{Z}_{h,d}|} \sum_{j \in \mathcal{Z}_{i,d}} M_{ij} \\ &= \frac{1}{n} \sum_{i \in \mathcal{X}} \sum_{d \in \Delta} \sum_{j \in \mathcal{Z}_{i,d}} M_{ij} \sum_{k \in \mathcal{Z}_{h,d}} \frac{1}{|\mathcal{Z}_{h,d}|}\end{aligned}$$

and now the summations over j can be joined together

$$= \frac{1}{n} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{Z}} M_{ij} = 1$$

Then, for the elements of the diagonal we have:

$$M'_{hh} = \frac{1}{n} \sum_{i \in \mathcal{X}} M_{ii}$$

So they are all the same, and it easily follows that $M'_{i,i} = \max_i^{M'}$ for all $i \in \mathcal{X}$, and that $I_\infty^M(X) = I_\infty^{M'}(X)$.

It remains to show that M' satisfies ϵ-differential privacy. We need to show that

$$M'_{hk} \leq e^\epsilon M'_{h'k} \qquad \forall h, h', k \in \mathcal{X} : d(h,h') = 1$$

From the triangular inequality we have (since $d(h,h') = 1$)

$$d(h',k) - 1 \leq d(h,k) \leq d(h',k) + 1$$

Thus, there are 3 possible cases: a) if $d(h,k) = d(h',k)$ the result is immediate since $M'_{hk} = M'_{h'k}$. b) Consider the case $d(h,k) = d(h',k) - 1$. We define

$$\mathcal{S}_{i,j} = \{j' \in \mathcal{Z} | j' \sim j \wedge d(i,j') = d(i,j) + 1\}$$

Note that $|\mathcal{S}_{i,j}| = (u - d(i,j))(v-1)$ (i and j are equal in $u - d(i,j)$ elements, we can change any of them in $v-1$ ways). The following inequalities hold:

$$\begin{aligned}
M_{ij} &\leq e^\epsilon M_{ij'} \qquad \forall j' \in \mathcal{S}_{i,j} \qquad \text{(diff. privacy)} \Rightarrow \\
(u - d(i,j))(v-1)M_{ij} &\leq e^\epsilon \sum_{j' \in \mathcal{S}_{i,j}} M_{ij'} \qquad \text{(add all the above)} \Rightarrow \\
\sum_{j \in \mathcal{Z}_{i,d(h,k)}} (u - d(h,k))(v-1)M_{ij} &\leq e^\epsilon \sum_{j \in \mathcal{Z}_{i,d(h,k)}} \sum_{j' \in \mathcal{S}_{i,j}} M_{ij'}
\end{aligned}$$

Let $d = d(h,k)$. Note that each $j' \in \mathcal{Z}_{i,d+1}$ is contained in exactly $d+1$ different sets $\mathcal{S}_{i,j}, j \in \mathcal{Z}_{i,d}$. So the right-hand side above sums all elements of $\mathcal{Z}_{i,d+1}$, $d+1$ times each. Thus we get

$$(u-d)(v-1) \sum_{j \in \mathcal{Z}_{i,d}} M_{ij} \leq e^\epsilon \ (d+1) \sum_{j \in \mathcal{Z}_{i,d+1}} M_{ij} \tag{9}$$

Finally, note that $|\mathcal{Z}_{h,d}| = \binom{u}{d}(v-1)^d$. We have

$$\begin{aligned}
M'_{hk} &= \frac{1}{n|\mathcal{Z}_{h,d}|} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{Z}_{i,d}} M_{ij} \\
&\leq e^\epsilon \ \frac{1}{n} \frac{d+1}{(u-d)(v-1)} \frac{1}{\binom{u}{d}(v-1)^d} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{Z}_{i,d+1}} M_{ij} \qquad \text{(from (9))} \\
&\leq e^\epsilon \ \frac{1}{n} \frac{1}{\binom{u}{d+1}(v-1)^{d+1}} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{Z}_{i,d+1}} M_{ij} \\
&= e^\epsilon M'_{h'k}
\end{aligned}$$

c) Symetrically for the case $d(h,k) = d(h',k) + 1$.

□

Lemma 3. *Let M be a channel matrix with dimensions $n \times m$ ($m \geq n$) from A to B. We fix the adjacency relation to the one defined in Section 3. Assume that there is an injective function $\zeta() : \mathcal{A} \to \mathcal{B}$ such that $M_{i,\zeta(i)} = \max^M_{\zeta(i)}$ for all $i \in \mathcal{A}$. Moreover, assume that there is an automorphism σ in $(\mathcal{A}, \sim)$ with only one orbit. If M satisfies ϵ-differential privacy then we can construct a new channel matrix M' such that:*

1. *M' respects ϵ-differential privacy;*
2. *$M'_{i,\zeta(i)} = M'_{h,\zeta(h)}$ for all $i, h \in \mathcal{X}$;*
3. *$M'_{i,\zeta(i)} = \max^{M'}_{\zeta(i)}$ for all $i \in \mathcal{X}$;*

4. $I_\infty^M(A) \leq I_\infty^{M'}(A)$.

Proof. For every $0 \leq h \leq n$ and $0 \leq k \leq m$ Let us define the elements of M' as:

$$M'_{hk} = \frac{1}{n} \sum_{i=0}^{n-1} M_{\sigma^i(h)\zeta(\sigma^i(k))} \tag{10}$$

First we prove that M' respects ϵ-differential privacy. For every pair $h \sim h'$ and every k:

$$\begin{aligned} M'_{hk} &= \sum_{i=0}^{n-1} M_{\sigma^i(h)\zeta(\sigma^i(k))} \\ &\leq \sum_{i=0}^{n-1} e^\epsilon M_{\sigma^i(h')k} \qquad \text{(by } \epsilon\text{-diff. privacy)} \\ &= e^\epsilon M'_{h'k} \end{aligned} \tag{11}$$

Now we prove that for every h, $M'_{h,\cdot}$ is a legal probability distribution.

$$\begin{aligned} \sum_{k=0}^{n-1} M'_{hk} &= \sum_{k=0}^{n-1} \frac{1}{n} \sum_{i=0}^{n-1} M_{\sigma^i(h)\zeta(\sigma^i(k))} \\ &= \sum_{i=0}^{n-1} \frac{1}{n} \sum_{k=0}^{n-1} M_{\sigma^i(h)\zeta(\sigma^i(k))} \\ &= \sum_{i=0}^{n-1} \frac{1}{n} \cdot 1 \\ &= 1 \end{aligned} \tag{12}$$

Next we prove that the maximum of every column is in the diagonal, i.e., for every k, $M'_{kk} = \max_k^{M'}$.

$$\begin{aligned} M'_{kk} &= \frac{1}{n} \sum_{k=0}^{n-1} M_{\sigma^i(k)\zeta(\sigma^i(k))} \\ &\geq \frac{1}{n} \sum_{k=0}^{n-1} M_{\sigma^i(h)\zeta(\sigma^i(k))} \\ &= M'_{hk} \end{aligned} \tag{13}$$

Finally, we prove that $I_\infty^{M'}(A) \geq I_\infty^M(A)$. It is enough to prove that $H_\infty^{M'}(A) \geq H_\infty^M(A)$.

$$
\begin{aligned}
H_\infty^{M'}(A) &= \sum_{h=0}^{n-1} M_{hh} \\
&\geq \sum_{h=0}^{n-1} \frac{1}{n} \sum_{i=0}^{n-1} M_{\sigma^i(h)\zeta(\sigma^i(k))} \text{ (since } M' \text{ has only one orbit)} \\
&= \sum_{h=0}^{n-1} \frac{1}{n} H_\infty^M(A) \\
&= H_\infty^M(A)
\end{aligned}
\tag{14}
$$

□

Theorem 1. If $\mathcal{K}$ provides ϵ-*differential privacy* then the leakage associated to $\mathcal{K}$ is bounded from above as follows:

$$
I_\infty(X;Z) \leq u \log_2 \frac{v\, e^\epsilon}{(v-1+e^\epsilon)}
$$

Proof. First we represent the randomized function $\mathcal{K}$ as a channel (X, Z, M) where:

- X is the input, whose career is the set of all possible databases $\mathcal{X} = V^u$;
- Z is the output, whose career is the set of all possible reported answers $\mathcal{Z}$;
- M is the conditional probabilities channel matrix $p_{Z|X}(\cdot|\cdot)$.

We assume without loss of generality that the matrix M has been obtained by taking the original matrix and then applying Lemma 1 to it, immediately followed by Lemma 3. Let us call a the value of every element in the diagonal of M.

Let us take an element $M_{i,i} = a$. For element in the border $Border(M_{i,i}, d)$ at distance d from $M_{i,i}$. The probability of this element can be at most $\frac{a}{e^{d\epsilon}}$. Also, the elements of row i represent a probability distribution, so they sum up to 1:

$$
\begin{aligned}
a + \sum_{d=1}^{u} \binom{u}{d} (v-1)^d \frac{a}{e^d \epsilon} = 1 \implies \\
\sum_{d=0}^{u} \binom{u}{d} (v-1)^d \frac{a}{e^{d\epsilon}} = 1 \implies \\
a \sum_{d=0}^{u} \binom{u}{d} (v-1)^d (e^\epsilon)^{u-d} = (e^\epsilon)^u \implies \\
a(v-1+e^\epsilon)^u = (e^\epsilon)^u \implies \\
a = \left(\frac{e^\epsilon}{v-1+e^\epsilon} \right)^u
\end{aligned}
\tag{15}
$$

Therefore:

$$\begin{aligned} I_\infty^M(X) &= H_\infty(X) - H_\infty^M(X) \qquad \text{(by definition)} \\ &= \log_2 v^u + \log_2 a \\ &= \log_2 v^u + \log_2 \left(\tfrac{e^\epsilon}{v-1+e^\epsilon}\right)^u \quad \text{(by (20))} \\ &= u \log_2 \tfrac{ve^\epsilon}{v-1+e^\epsilon} \end{aligned} \tag{16}$$

□

Proposition 1. For every u, v, and ϵ there exists a randomized function $\mathcal{K}$ which provides ϵ-differential privacy and whose leakage, for the uniform input distribution, is $I_\infty(X;Z) = B(u,v,\epsilon)$.

Proof. The adjacency relation in $\mathcal{X}$ determines a graph structure $G_\mathcal{X}$. Set $\mathcal{Z} = \mathcal{X}$ and define the matrix of $\mathcal{K}$ as follows:

$$p_\mathcal{K}(z|x) = \frac{B(u,v,\epsilon)}{(e^\epsilon)^d} \qquad \text{where } d \text{ is the distance between } x \text{ and } z \text{ in } G_\mathcal{X}$$

It is easy to see that $p_\mathcal{K}(\cdot|x)$ is a probability distribution for every x, that $\mathcal{K}$ provides ϵ-differential privacy, and that $I_\infty(X;Z) = B(u,v,\epsilon)$. □

Lemma 4. *Let Y and Z be two random variables with domains $\mathcal{Y}$ and $\mathcal{Z}$, respectively. If a randomization mechanism $\mathcal{H} : Y \to Z$ respects an ϵ-ratio in the sense that $p_\mathcal{H}(z|y') \le e^\epsilon \cdot p_\mathcal{H}(z|y'')$ for all $y', y'' \in \mathcal{Y}$ and $z \in \mathcal{Z}$, then the Rényi-min mutual information between Y and Z is bounded by:*

$$I_\infty(Y;Z) \le \epsilon \log e$$

Proof. For clarity reasons, in this proof we use the notation $p(z|Y=y)$ for the probability distributions $p_\mathcal{H}(z|Y=y)$ induced by the mechanism $\mathcal{H}$.

Let us calculate the Rényi mutual information using the formula $I_\infty(Y;Z) = H_\infty(Y) - H_\infty(Y|Z)$.

$$
\begin{aligned}
-H_\infty(Y|Z) &= \log \sum_z p(z) \max_y p(y|z) && \text{(by definition)} \\
&= \log \sum_z \max_y p(z)p(y|z) \\
&= \log \sum_z \max_y p(y)p(z|y) && \text{(by probability laws)} \\
&\leq \log \sum_z \max_y p(y)e^\epsilon p(z|\hat{y}) && \text{(for any fixed } \hat{y} \in \mathcal{Y}\text{, since} \\
& && \mathcal{H} \text{ satisfies } \epsilon\text{-ratio)} \\
&= \log \sum_z e^\epsilon p(z|\hat{y}) \max_y p(y) \\
&= \log \left(e^\epsilon \max_y p(y) \sum_z p(z|\hat{y}) \right) \\
&= \log \left(e^\epsilon \max_y p(y) \right) && \text{(by probability laws)} \\
&= \log e^\epsilon + \log \max_y p(y) \\
&= \epsilon \log e - H_\infty(Y) && \text{(by definition)}
\end{aligned} \tag{17}
$$

Therefore:

$$
H_\infty(Y|Z) \geq H_{\infty(Y)} - \epsilon \log e \tag{18}
$$

And it follows that:

$$
\begin{aligned}
I_\infty(Y;Z) &= H_\infty(Y) - H_\infty(Y|Z) && \text{(by definition)} \\
&\leq H_\infty(Y) - H_\infty(Y) + \epsilon \log e && \text{(by Equation 18)} \\
&= \epsilon \log e
\end{aligned} \tag{19}
$$

□

Theorem 2. If $\mathcal{K}$ provides ϵ-*differential privacy* then for all $D^- \in Val^{u-1}$ the leakage about an individual is bounded from above as follows:

$$
I_\infty(X_{D^-}; Z) \leq \log_2 e^\epsilon
$$

Proof. By Lemma 4, an oblivious randomization mechanism $\mathcal{H}$ with input Y and output Z has the bound $I_\infty(Y;Z) \leq \epsilon \log e$. However, by the information processing inequality, $I_\infty(X;Z) \leq I_\infty(Y;Z)$ and the theorem follows if we take X to be Val^{u-1}. □

Theorem 3. Let $\mathcal{H}$ be a randomization mechanism for the randomized function $\mathcal{K}$ and the query f, and assume that $\mathcal{K}$ provides ϵ-differential privacy. Assume that $(\mathcal{Y}, \sim)$ admits a graph automorphism with only one orbit. Furthermore, assume that there exists a natural number c and an element $y \in \mathcal{Y}$ such that, for every d, either $|Border(y,d)| = 0$ or $|Border(y,d)| \geq c$. Then

$$
\mathcal{U}(X,Y) \leq \frac{(e^\epsilon)^n(1-e^\epsilon)}{(e^\epsilon)^n(1-e^\epsilon) + c\,(1-(e^\epsilon)^n)}
$$

Proof. First we represent the randomized function $\mathcal{K}$ as a channel (Y, Z, M) where:

- Y is the input with carrier $\mathcal{Y}$ representing all possible real answers for the query;
- Z is the output with carrier $\mathcal{Z}$ representing all possible reported answers for the query;
- M is the conditional probabilities channel matrix $p_{Z|Y}(\cdot|\cdot)$.

We assume without loss of generality that the matrix M has been obtained by taking the original matrix and then applying Lemma 1 to it, immediately followed by Lemma 3. Let us call a the value of every element in the diagonal of M.

Let us take an element $M_{i,i} = a$. For element in the border $Border(M_{i,i}, d)$ at distance d from $M_{i,i}$. The probability of this element can be at most $\frac{a}{e^{d\epsilon}}$. Also, the elements of row i represent a probability distribution, so they sum up to 1:

$$
\begin{aligned}
a + \sum_{d=1}^{n} |Border(y,d)| \frac{a}{(e^\epsilon)^d} = 1 \implies & \text{(since by hypothesis } |Border(y,d)| \geq c) \\
a + \sum_{d=1}^{n} c \frac{a}{(e^\epsilon)^d} \leq 1 \implies & \\
a\,(e^\epsilon)^n + \sum_{d=1}^{n} c\,(e^\epsilon)^{n-d} \leq (e^\epsilon)^n \implies & \\
a\,(e^\epsilon)^n + \sum_{t=0}^{n-1} c\,(e^\epsilon)^t \leq (e^\epsilon)^n \implies & \text{(geometric progression sum)} \\
a\,(e^\epsilon)^n + c\,a\,\frac{1-(e^\epsilon)^n}{1-e^\epsilon} \leq (e^\epsilon)^n \implies & \\
a \leq \frac{(e^\epsilon)^n(1-e^\epsilon)}{(e^\epsilon)^n(1-e^\epsilon) + c\,(1-(e^\epsilon)^n)} & \qquad (20)
\end{aligned}
$$

Since $\mathcal{U}(Y, Z) = a$, the theorem follows.

□

Theorem 4. The definition in (6) determines a legal channel matrix for $\mathcal{H}$, i.e., for each y, $p_{Z|Y}(\cdot|y)$ is a probability distribution. Furthermore, it meets the differential privacy requirement, in the sense that the the composition $\mathcal{K}$ of f and $\mathcal{H}$ provides ϵ-differential privacy. Finally, $\mathcal{H}$ is optimal in the sense that it maximizes utility when the input distribution (i.e. the distribution of Y) is uniform.

Proof. We follow a reasoning analogous to the proof of Theorem 3, but using $|Border(y,d)| = c$, to prove that $\mathcal{U}(Y, Z) = \frac{(e^\epsilon)^n(1-e^\epsilon)}{(e^\epsilon)^n(1-e^\epsilon)+c\,(1-(e^\epsilon)^n)}$. From the same theorem, we know that this is a maximum for the utility. □