Onto-DP: Constructing neighborhoods for differential privacy on ontological databases

Yasmine Hayder ✉, Adrien Boiret, Cédric Eichler[0000−0003−3026−1749], and Benjamin Nguyen

LIFO, INSA CVL, Univ. Orléans, Inria, France
firstname.lastname@insa-cvl.fr

Abstract. In this paper, we investigate how attackers can discover sensitive information embedded within databases by exploiting inference rules. We demonstrate the inadequacy of naively applied existing state of the art differential privacy (DP) models in safeguarding against such attacks.
We introduce ontology aware differential privacy (Onto-DP), a novel extension of differential privacy paradigms built on top of any classical DP model by enriching it with semantic awareness. We show that this extension is a sufficient condition to adequately protect against attackers aware of inference rules.

Keywords: Privacy · Differential privacy · Inferences · Ontology.

1 Introduction

Databases in general, and semantic databases such as Knowledge Graphs (KGs) in particular, are often used to store personal and/or private information, such as healthcare data [1]. In this article, we are interested in the *privacy* issues linked to protecting the personal information in databases when publishing the results of queries. These privacy questions have been a pressing issue since seminal works of Sweeney [2] on k-anonymity, and a whole field called *differential privacy* (DP) [3] has seen a very strong development these last twenty years. DP has garnered a great interest within both theoretical and applied database and privacy communities. The general idea behind DP is to bound the relative information gain on the underlying data that a querier obtains when observing the result of their query. One of the most important features of DP is that the approach (supposedly) makes no assumption on any background knowledge that the attacker may possess, which was an important breakthrough, compared to previous works such as k-anonymity, where a large part of the practical difficulties when trying to evaluate the security of the approach is that it depends on the attacker's background knowledge. However there is in fact an implicit dependency of DP on some kind of attacker background knowledge: DP is based on the concept of *adjacent* databases, loosely defined as two databases that differ by "one element". In this article, we consider that the adjacent databases of a

given database D are given by defining a distance d and its associated *neighborhood* $N_d(D)$. Once this neighborhood definition is given, it is then possible to directly apply state of the art DP mechanisms to protect the data, such as adding random noise drawn from a *Laplace distribution*.

A lot of research in DP has gone into building optimal mechanisms, however we focus on a much less studied question: *how to correctly define this distance*, in order to capture what we call *semantic attackers* that possess knowledge about inference rules (Data dependency). To our knowledge, we are the first to (correctly) build such a distance. The main difficulty of this research question is that classical distances used in DP are expressed over a representation-dependent notion of single datum (e.g. relational databases differ by a single tuple, graph databases differ by a single node or edge, RDF databases differ by a single triple, etc.), but this may not precisely translate two databases differing in one fact, in the semantic sense, since data items are often correlated, especially in graph structured databases. Works such as Pufferfish [4] provide a very wide setting to customize privacy for dependent data by defining secrets, alternative worlds that must be indistinguishable, and an attacker's background knowledge. However, while DP and deterministic inference considerations can be expressed in this setting with a prohibitively comprehensive class of attackers, in practice showing that a process respects a privacy constraint or not is best checked on a small set of worst-case-scenario attackers. No such suitable definition that we know of exist for databases with inference.

To address these limitations, we (i) formalize both the concepts of a *semantically aware attacker*, *i.e.* one that knows how to infer facts from existing information, and knows all but one information, and of a distance providing a *well suited defense* w.r.t. a class of attacker, in the context of a DP based protection; (ii) demonstrate that classical DP approaches on databases, while *well suited* w.r.t. semantic-unaware attackers, are *ill suited* w.r.t. attackers aware of inference rules; (iii) show that such ill-suitedness can lead to incorrect estimation of privacy and leaks of supposedly protected data; (iv) propose *onto-DP*, a consistent extension built on top of existing DP models, and demonstrate that it is (by construction) *well suited* for semantic-aware attackers

The remainder of this paper is structured as follows. Sec. 2 introduces the background on DP and (semantic) databases. The problem we study is presented and formalized in Sec. 3. The incorporation of inference rules is discussed in Sec. 4. The related work and positioning of our paper are presented in Sec. 6. Finally, Sec. 7 concludes and presents future research directions.

2 Background and notations

In this section, we provide the background on the data model, inference rules, differential privacy and distances necessary to the theoretical foundations of this paper.

Private database D. We consider a database D containing (sensitive) information to be queried. We do not make any extra hypothesis on the structure of

D: it may be relational, a knowledge graph, etc. We note $\mathcal{D}$ the set of databases considered and $D \subseteq D'$ the fact that D is contained in D'.

As we consider attackers that conduct inferences, we will instantiate D as a knowledge graph (KG) in our examples, but our results hold in the general case. Knowledge graphs are structured representation in graphs that model real-world entities as nodes, and their relationships as edges, or subject-predicate-object triples. RDF or Neo4j [5] are common frameworks to represent KGs.

Example 1 (Hospital DB). A toy example knowledge graph database

```
# Example Hospital Database
@prefix ex:  <http://example.com/hospital#> .
ex:d1  a       ex:doctor ;
       ex:worksIn ex:dept1 ;
       ex:hasPatient ex:p1 .
ex:dpt1 a      ex:dept .
ex:p1  a       ex:patient .
```

Inference Rules I are mechanisms that allow the derivation of new knowledge from existing facts by applying logical reasoning to help enrich the database with new facts. They may be expressed in various knowledge representation languages, from the simplest ones like RDFS [6] to much more expressive languages such as SWRL [7]. Reasoner engines such as Hermit [8] are used to derive new facts in the database, based on its contents and the rules considered.

Example 2 (Inferring new tuples). Consider an inference rule stating that a patient under the care of a physician working in a particular department is a patient in said department, written in SWRL "*human readable syntax*" format [7]:

$$IR = \{\texttt{hasPatient(?x, ?y)} \wedge \texttt{worksIn(?x, ?z)} \Rightarrow \texttt{patientIn(?y, ?z)}\}$$

We show next the hospital database of Example 1, after applying a reasoner.

```
# Example Hospital Database after reasoning
@prefix ex:  <http://example.com/hospital#> .
ex:d1  a       ex:doctor ;
       ex:worksIn ex:dept1 ;
       ex:hasPatient ex:p1 .
ex:dpt1 a      ex:dept .
ex:p1  a       ex:patient .
ex:p1  ex:patientIn ex:dept1 .
```

We formally define the inference system by the following:

Definition 1 (Inference System). *Let $\mathcal{D}$ be a space of databases, and I a set of inference rules. An inference system $I : \mathcal{D} \to \mathcal{D}$ is function that associates some database D of $\mathcal{D}$ to its* saturated *version $I(D)$ which is obtained by applying all the rules in I. It is idempotent $I(I(D)) = I(D)$.*

Notations: We note $I = \emptyset$ to indicate that *there are no inference rules, thus no extra information may be inferred.* Formally, this means I is the identity function. We say that a database D is a *saturation* of $D^{(-1)}$ by inference system I if $I(D^{(-1)}) = D$. Conversely, we say that $D^{(-1)}$ is an antecedent of D if D is a saturation of $D^{(-1)}$. Note that the database in Example 2 is indeed saturated by the inference system I, applying the inference rule IR.

Differential Privacy. DP [3] is possibly the most well-established criterion in the privacy research community. It ensures that an attacker who observes the outcome of a query cannot infer the presence or absence (and hence the value) of any particular sensitive datum in the dataset.

Definition 2 (ε-Differential privacy DP [3]). *Given $\varepsilon > 0$, a function f defined on $\mathcal{D}$ and a distance d over $\mathcal{D}$, f satisfies ε-DP if for all $(D_1, D_2) \in \mathcal{D}^2$ such that $d(D_1, D_2) = 1$, and for all subsets S of the range of f, we have:*

$$\Pr[M(D_1) \in S] \leq e^{\varepsilon} \times \Pr[M(D_2) \in S]$$

where probability is taken over the randomness of f. In this case, we say that D_1 and D_2 are ε-indistinguishable [9] where ε represents the privacy budget and parameterizes the protection.

Implementing DP. A classical method to implement a DP mechanism for numeric queries is to return a noisy answer rather than the true query result [10]. The added noise must be carefully calibrated so that the result remains useful while still protecting individual contributions. Its amplitude depends on ε and the query's sensitivity Δf (called ℓ_1-sensitivity by [10]), i.e. how much it may vary among a neighborhood:

Definition 3 (Global Sensitivity Δf [10]). *Given a numeric query $f : \mathcal{D} \to \mathbb{R}$ and a distance d over $\mathcal{D}$, the (global or ℓ_1) sensitivity of f is defined as*
$\Delta f = \max_{x,y} |f(x) - f(y)|$ *for all x, y such that $d(x, y) = 1$.*

DP on various databases models. DP is immediately applicable to any space $\mathcal{D}$ given a proper distance d or simply a neighborhood definition. Classically, for relational databases, the notion of neighborhood corresponds to two databases D_1 and D_2 differing by a single tuple. When considering graph databases, two DP-models, relying on two notions of neighborhoods, prevail: k-edge-DP [11] and node-DP [12], where two databases are neighbors if they differ by up to k edges (resp. one node and all its adjacent edges).

Bounded and Unbounded DP. There exist two ways to compute distances for differential privacy: unbounded and bounded DP [13]. In bounded-DP, two datasets are considered neighbors if one can be obtained by editing/modifying one sensitive piece of information within the other. In unbounded-DP, two datasets are neighbors if they differ by the addition or deletion of a single piece of information.
We now define *paired* bounded and unbounded distances, an important new concept that we introduce to link the attacker/defender models:

Definition 4 (Paired bounded/unbounded distances). *Let D_1 and D_2 represent two databases. Two distances d_u d_b, unbounded and bounded, respectively, are said to be paired if :*

$$\forall D_1, D_2 | D_1 \neq D_2, d_b(D, D') = 1 \iff \exists D_0 | D_0 \subseteq D_1 \wedge D_0 \subseteq D_2 \\ \wedge\, d_u(D_0, D_1) = 1 \wedge d_u(D_0, D_2) = 1$$

3 Problem formalization and analysis

In this paper, we consider a curator (defender) relying on DP where the true database D is ε-indistinguishable from its neighbors according to some distance d. On the other hand, the attacker tries to determine the true database among a set of databases it considers possible.

Our attackers generalize classical worst-case scenario attackers that are just one datum (e.g. edge or line) away from knowing the full graph, to a setting where fact inference is possible. This means that the attacker is considering *unbounded* distance neighbors where exactly one piece of information ι_0 (i.e. the missing datum and all other derived from it) is added to its prior knowledge. On the other hand, the curator (who knows D) must protect against *any* such attacker (for all possible values of ι_0). The ε-indistinguishable databases should hence be the union of all the databases considered by *any* attacker, thus the union of all the databases containing all information of D except for one plus all possible variations of the missing information. This is by construction *exactly* the set of databases that are *bounded* neighbors of D. We will thus consider any paired bounded/unbounded distances d_b and d_u, with d_b the distance used by the *curator* (adding the noise) and d_u used to characterize the attacker.

In this section, we formalize the attackers, the defense and attack spaces, and the concept of well-suitedness. By using these definitions, we are able to show that in the classical case with no consideration of inferences, existing DP models are well suited for our attackers. *An interesting problem, studied in the rest of the paper, arises when the attacker leverages knowledge of inference rules.* In this case, classical DP models are no longer well-suited. The main result of this paper is to restore this property of well-suitedness with an extension of classical DP models that applies for various distances accounting for ontologies.

Defense space. A defense space is a mapping from $\mathcal{D}$ to $2^{(\mathcal{D})}$ that maps each database to a set of decoys. In DP, this set of decoys (and their corresponding defense space) is the neighborhood of the true database according to the chosen distance (plus the original database itself).

Definition 5 (Defense space). *Let d_b be a (bounded) distance over databases. The defense space N_{d_b} of d_b maps each graph D to $\{D' | d_b(D, D') \leq 1\}$.*

Attacker model. An attacker is an observer that starts with some knowledge of the database (i.e. a subset of data that they know to be true). This allows them to construct a set of databases they believe are susceptible to be the true state

of the database, which we call *attack space*. We consider a *worst-case* (i.e. very knowledgeable) attacker that has knowledge of the database *up to exactly one* missing datum, as in [14,15]. Such an attacker (that we term *up-to-one* attacker) and its attack space are defined as follows:

Definition 6 (Up-to-one attacker and its attack space). *Let d_u be an (unbounded) distance on $\mathcal{D}$, I an inference system, D_0 be a database modeling the prior knowledge of the attacker. We note $A^I_{d_u}(D_0)$ the up-to-one attacker on distance d_u, aware of inference system I, and of prior D_0. $A^I_{d_u}(D_0)$ considers all saturated graphs $D' \in \mathcal{D}$ such that $\exists D'^{(-1)}$ such that*

- $D' = I(D'^{(-1)})$
- $D_0 \subseteq D'^{(-1)}$
- $d_u(D_0, D'^{(-1)}) = 1$

Intuitively, such an attacker has a prior knowledge D_0 and misses a single datum of information. The attacker considers all databases $D'^{(-1)}$ with exactly one possibility for the datum (i.e. $d_u(D_0, D'^{(-1)}) = 1$) and saturates them using the inferences I. By notational abuse, $A^I_{d_u}(D_0)$ denotes both an attacker and its attack space and $A^I_{d_u}$ denotes both the class of attackers and the union of their attack spaces.

Well-suitedness of defense with regard to an attacker. A defense space is appropriately calibrated against a class of attackers if the set of graphs they considered plausible is exactly equal to the defense space. If we consider DP, and the up-to-one attackers, we get the following formalization:

Definition 7 (Well-suited DP defense). *Let d_u, d_b be two distances on $\mathcal{D}$ and inference system I. We say that d_b-DP is a well-suited defense to the $A^I_{d_u}$ up-to-one class of attackers if, for all $D \in \mathcal{D}$,*

$$N_{d_b}(D) = \bigcup_{D_0 | D \in A^I_{d_u}(D_0)} (A^I_{d_u}(D_0))$$

We note that this property is indeed respected for classical DP and distance in the absence of inference rules.

Lemma 1 (Well suitedness of classical DP model w.r.t. semantic unaware attacker). *For d_u and d_b paired distances, d_b-DP is a well-suited defense to the $A^{\emptyset}_{d_u}$ up-to-one class of attackers.*

Objectives and contributions of the rest of the paper. In the next part of the paper, we analyze the ill-suitedness (and related ill-effects) of classical DP models w.r.t. semantic aware attackers and propose (I, d_b)-DP that integrates ontology to a distance d_b, such that:

1. $(\emptyset, d_b)$-DP is equivalent to d_b-DP
2. if d_b, d_u are paired with d_b bounded, then (I, d_b)-DP is well suited w.r.t. $A^I_{d_u}$.

4 Mismatch of the $(\emptyset, d_b)$-DP defense for a semantic aware attacker $A^I_{d_u}$

We now study the impact of the up-to-one adversary knowing inference rules and show that classical distance are ill suited against such an attacker.

A curator using $(\emptyset, d_b)$-DP associates to each database D the defense space $N_{(\emptyset,d_b)}(D) = \{D' | d_b(D, D') \leq 1\}$. Given a database D, by Def 6, an instance of $A^I_{d_u}$ considering D plausible starts with a prior D_0 such that $\exists D^{(-1)}, D_0 \subseteq D^{(-1)} \wedge d_u(D^{(-1)}, D_0) = 1 \wedge I(D^{(-1)}) = D$. It only considers other databases D' with an antecedent $D'^{(-1)}$ such that $d_u(D'^{(-1)}, D_0) = 1$. The problem arises here since it is possible to have $d_b(D, D') > 1$ even though $d_b(D^{(-1)}, D'^{(-1)}) = 1$. Databases an attacker with inference rules considers are not necessarily neighbors in the sense of the considered distances d_u or d_b. Indeed, inferences on different d_b-neighboring databases can create (arbitrarily) distant databases w.r.t. d_b, since an arbitrary number of facts could be added during the saturation process (this obviously depends on I and D).

Since the defense space of the curator is entirely composed of d_b neighbors of D, it immediately follows that there may be a mismatch and that the $(\emptyset, d_b)$-DP defense cannot always be well-suited for a semantic aware attacker $A^I_{d_u}$. This is particularly problematic since the perceived query variation of an up-to-one attacker may be *greater* than the curator's considered sensitivity. For any query Q applicable on $\mathcal{D}$, we note $\Delta_I Q$ the perceived sensitivity of $A^I_{d_u}$ attackers, i.e.

$$\Delta_I Q = \max_{D_0}\Big(\max_{(D,D')\in(A^I_{d_u}(D_0))^2} |Q(D) - Q(D')|\Big)$$

Proposition 1 (Privacy leakage by a $(\emptyset, d_b)$ curator against an $A^I_{d_u}$ attacker). *Consider $A^I_{d_u}$, a class of attackers with knowledge of inference rules on a database, and a curator using $(\emptyset, d_b)$-DP, thus a defense space not considering inference rules. It is possible for such a curator to underestimate the leakage of sensitive private information; i.e. that for a query Q, $\Delta_\emptyset Q < \Delta_I Q$. In fact, it is possible that $\Delta_\emptyset Q = 0$ while $\Delta_I Q > 0$.*

Proof. We illustrate this mismatch in an example inspired by [16]. We consider edge-distances and $\mathcal{D}$ to be KGs such as the one in Fig. 1 representing de-identified data in a hospital. In such graphs, each d_i is *necessarily* assigned to a $dept_j$, and each p_k is *necessarily* a patient of a specific d_i. All of them are instances of a class `dept`, `doctor`, or `patient`, which are represented as nodes. Typing is represented as an edge. Edges are labeled *hasPatient*, *worksIn*, *hasType*, or *patientIn*. Contrarily to aformentioned relations, the *patientIn* relation is *not mandatory* and links the `patient` of a `doctor` to the `departement` it belongs to.

We consider the inference rule stated in Example 2 that a `patient` under the care of a `doctor` working in a particular `dept` is a `patient` in said `dept`.

We now consider an example database D^{ex} and its saturation by I, as represented in Figure 2, where types are implicit for simplicity sake. We consider the basic query Q which counts the number of patients in the `oncology` department.

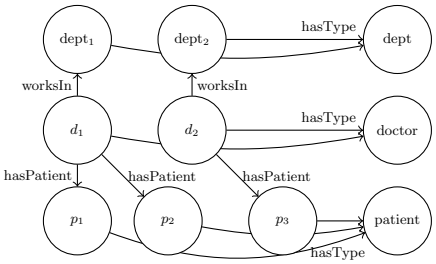


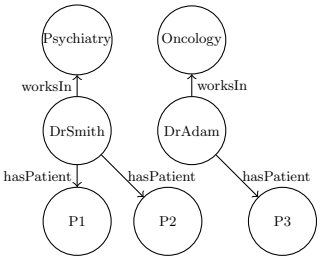
Fig. 1: Considered databases $\mathcal{D}$

```
Q = SELECT (COUNT(DISTINCT ?patient) AS ?numPatients)
WHERE {
  ?patient :patientIn :Oncology .
}
```

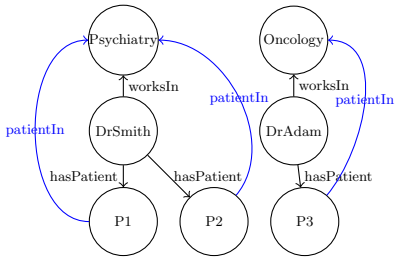
We further restrict $\mathcal{D}$ to only contain saturated database, considering that the curator systematically saturates the database as Q would e.g. returns 0 on D^{ex}.

If an attacker can start from a prior D_0, guess a database D' with one more piece of data (*i.e.* one more edge), then saturate that D' into a database $I(D') \in \mathcal{D}$, then $I(D')$ will be considered by the attacker as part of the attack space. *This means that an attacker considers a specific saturated database if its prior is a subgraph neighbor of one of this database's antecedents*, i.e. a graph that saturates to it. In order to provide adequate protection, the curator should also consider them as part of the defense space.

In Fig. 3a, database D_0^{ex} represents the attacker's knowledge. Note that in this case, $D_0^{ex} \notin \mathcal{D}$ since there exists a *doctor* that does not work in any *department*.

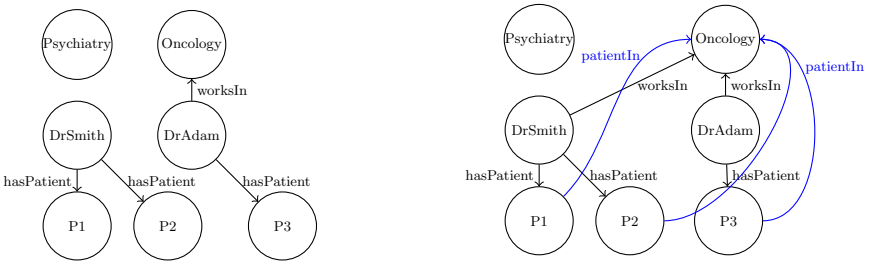


(a) Antecedent of the true database D^{ex}



(b) True database $I(D^{ex})$

Fig. 2: True database and inferred information



(a) Attacker's prior knowledge D_0^{ex}

(b) Plausible graph for $A^{C,I}_{d_u^e}(D_0^{ex})$

Fig. 3: Example of attacker knowledge and saturation of a possible neighbor

The attacker knows $I = \{\texttt{hasPatient(?x, ?y)} \wedge \texttt{worksIn(?x, ?z)} \Rightarrow \texttt{patientIn(?y, ?z)}\}$, considers the neighbors of D_0^{ex}, then saturates them. Fig. 2b and Fig. 3b show two plausible saturated graphs for this attacker. This means that *by construction* they differ by a single datum of information for the attacker. Hence, $\Delta_I Q \geq 2$. However, from the curator's point of view, if they were to use $(\emptyset, d_b)$-DP as a defense without further adjustment, the database pictured in Fig. 3b) would not be part of the defense space of the database $I(D_0^{ex})$ (Fig. 2b), as they are too far from each other (the edge distance between these graphs is 3, not 1). This means that these graphs will not be considered by the curator when computing the sensitivity of the query.

Indeed, the direct $(\emptyset, d_b)$-distance neighbors of the database $I(D^{ex})$ (and actually many saturated graphs) will have the same answer to our query, as any alteration to a `worksIn` or `hasPatient` edge may have a knockdown effect on `patientIn` and would result in the database no longer be saturated (and hence $\notin \mathcal{D}$). In fact, two saturated graphs can only be $(\emptyset, d_b)$-distance neighbors if their only difference is swapping a patient between two doctors of same department, or changing the department of a patientless doctor. Under those circumstances, as stated in the proposition, we find a sensitivity equal to 0 under bounded edge DP on $\mathcal{D}$.

The code and demo of our tool illustrating this possible mismatch are available at: https://anonymous.4open.science/r/Onto-Differential-Privacy-E068/README.md

Consequence of this mismatch. In this case, this ill-suitedness has *drastic* consequences in particular when restricting the considered space to saturated databases: following the formal definition of DP, a sensitivity equal to 0 would mean that a query with no added randomized mechanism (i.e. outputing the true result of the query) is ε DP for all $\varepsilon > 0$. Indeed, since the result of the query does not vary among any neighborhood, the probability of obtaining a result is the same when applied to any adjacent databases. However *in reality*, the query's raw results *do* reveal some sensitive information to our up-to-one attacker, which could derive its missing piece of information *with absolute certainty*. The mismatch of attack and defense spaces would thus here lead to a dramatic **complete absence of protection under classical DP**.

5 (I, d_b)-onto-DP : A defense model against a semantic-aware attacker $A^I_{d_u}$

In order to solve the sensitivity issue when using traditional DP when dealing with ontology-aware attackers, we introduce (I, d_b)-onto-DP, based on a novel distance notion that contains a reasoning process. This distance is constructed to have a defense space matching semantic aware attackers. As such, it considers saturated graphs, and neighbors in this distance are not neighbors in the sense of d_b, but rather have unsaturated ancestors that are d_b neighbors. Hence, they are also unbounded neighbors of a common graph, matching with some attacker's prior. This is illustrated in Fig. 4 and formalized as follows.

Definition 8 ((I, d_b)-onto Distance). *Let d_b be a (bounded) distance. The (I, d_b)-ontology aware distance is defined by its neighborhoods, then classically extended. For $(D, D') \in \mathcal{D}^2$, D' is a neighbor of D if and only if there exists an antecedent $D^{(-1)}$ of D and $D'^{(-1)}$ of D' w.r.t. I such that $d_b(D^{(-1)}, D'^{(-1)}) = 1$.*

Consequence. It is immediate that $(\emptyset, d_b)$-ontology aware distance is d_b.

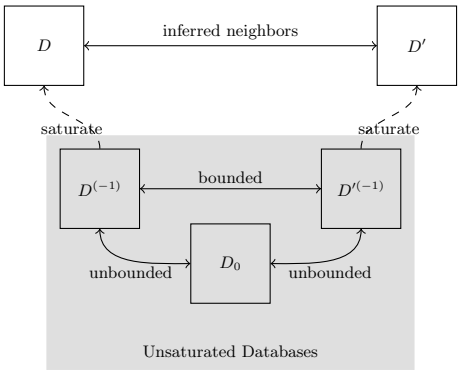


Fig. 4: (I, d) neighborhood pattern

Theorem 1 (Onto-DP is well-suited w.r.t. semantic aware attacker). *Let d_u, d_b be paired distances, unbounded and bounded respectively. For any I inference rules, (I, d_b)-DP is a well-suited defense to the $A^I_{d_u}$ up-to-one class of attackers*

Proof. Let us consider a database $D \in I(\mathcal{D})$, and a $A^I_{d_u}$ up-to-one attacker of prior D_0 that considers it. This means that there exists a database $D^{(-1)}$, neighbor of D_0 according to d_u, such that $I(D^{(-1)}) = D$. Any other databases considered by this attacker are $D' \in I(\mathcal{D})$ such that there exists a database $D'^{(-1)}$, d_u neighbor of D_0, whose saturated is $I(D'^{(-1)}) = D'$. Since $D^{(-1)}$ and

$D'^{(-1)}$ are both d_u neighbors of the same D_0, they are d_b neighbors by definition of paired distance. By definition, D and D' are neighbors in the (I, d_u)-distance.

Conversely, let us consider a database $D, D' \in I(\mathcal{D})$, neighbors in the (I, d_b)-distance. By definition, there exist $D^{(-1)}$ antecedent of D and $D'^{(-1)}$ of D' that are d_b neighbors. Since d_b and d_u are paired, this means that there exists some D_0, subset and d_u-neighbour of both $D^{(-1)}$ and $D'^{(-1)}$. This means that a (I, d_u) up-to-one attacker of prior D_0 considers both D and D'.

From both of these we conclude that the defense space of D in the (I, d_b)-onto-DP is exactly the set of all databases considered by at least one $A^I_{d_u}$ up-to-one attacker that also considers D.

Takeaway. Theorem 1 is the main result of the article. It shows that it is possible to build a DP mechanism that will correctly evaluate the sensitivity of queries, in the presence of an attackers that have knowledge of inference rules on a database, by using our proposed (I, d_b)-distance.

6 Related Work

Since KGs are the traditional representation for knowledge centered databases, we provide herein an overview of existing work related to semantic aware DP in such context. We also discuss work questioning attacks models for DP and investigating distances and neighborhoods for DP. Finally, we present the proposal closest to our own, DP approaches over correlated data.

DP for KGs. Even though KGs are the traditional representation for knowledge centered databases, there is surprisingly little work proposing semantic-aware DP approaches on KG. Standard DP approaches are oftentimes applied, for example, [17] applies "triple"-DP, which is equivalent to traditional edge-DP, to the problem of Federated Knowledge Graph Embedding.

Reuben [18] was, to the best of our knowledge, the first to propose semantic-aware DP for edge-labeled directed graphs. This approach was limited to applying edge-DP on a subset of labels. Building on this work, Taki et al. [19] proposed a projection-based approach to reduce the sensitivity of queries on KG using QL-edge-DP. Han et al. [20] proposed a similar idea where a set of sensitive relationships is specified.
To the best of our knowledge, semantic-aware DP approaches for KG are limited to the consideration of sensitive and non-sensitive labels (or types). There exists no work integrating constraints and inferences in DP approaches.

Attacks models on DP. The efficiency of the protection of DP is reliant on the choice of an ε. However, the concrete guarantees such constraints provide is heavily reliant on the type of attack scenario considered. In domains like membership or inference attacks [14], one standard scenario is the *informed attacker*, who knows everything about a model and its training data save for one specific element. This worst-case scenario is inspired by the implicit threat model of DP [21]. Indeed, previous works [15] use worst-case attack scenarii to measure

the efficiency of ε-DP. Their illustrative choice is that of an attacker that tries to identify members of a queriable subgroup of individuals, but only misses one information to do so, a scenario which serves as an inspiration for our up-to-one attacker. Some works like [22,23] explored relaxations of this worst-case attacker, but [24] demonstrates that such relaxations can lead to vulnerabilities.

Questioning DP-distances and Neighbourhoods. Several existing works also study the necessary departures from classic distances and mechanisms in scenarii where the neighborhoods given by classic distances are of variable believability or usefulness. Geo-indistinguishability [25] (and more generally metric-DP [26]), which aims to publish someone's location with enough noise so that an attacker cannot be certain of the person's true position, some locations (e.g. a river, the sea) are considered unlikely answers, that cannot realistically count as convincing decoys.

DP over correlated data. In many real cases like social networks, data is related and cannot be considered independent [27]. To deal with this, models like Pufferfish [4] and Bayesian Differential Privacy (BDP) [28] were developed. These formalisms provide a very wide setting to customize privacy for dependent data by defining secrets, *i.e.* alternative worlds that must be indistinguishable, and an attacker's background knowledge. Under such a setting, differential privacy can be defined as requiring any attacker, regardless of prior knowledge, to be unable to gain intel from an output. To add deterministic inference constraints is then to prune this large set of background knowledges by only considering those that never consider impossible databases where inferences are not properly made. While this accurately translates the definition of DP, and lets the authors show that under any correlation, classical Laplace mechanisms allow for security leaks, one of the keys to differential privacy's ease of use is that rather than checking for all attackers whether they can gain undue confidence on a specific secret, one can focus on showing it for worst-case scenario attackers, that only lack one element of a database, and extending the property to the general case. To find proper background knowledge for attackers that are able to use inference to deduce several facts from one guess is then necessary to have a notion of semantically aware DP one can hope to demonstrate on a given process.

7 Conclusion

This paper explores the challenges of protecting sensitive information against attackers with knowledge about the database semantics (Data dependency). We introduced and investigated *semantic aware* attackers, who have knowledge of the database related inference rules, and showed that traditional DP methods may underestimate the knowledge gained by an attacker, leading to privacy leaks. Thus, we provide tight theoretical bound on the involved (perceived) sensitivities. To address these issues, we proposed *onto-DP*, an extension of existing differential privacy paradigms that enrich them with the consideration of the database inference rules.

Future work. We believe these results open exciting new research directions at the intersection of DP and semantically rich databases such as KGs. A very promising direction is to extend this work beyond count queries, using e.g. the Exponential Mechanism [29] which is adapted for categorical queries, but that also relies on some way of measuring the distance between plausible answers, and thus where an investigation of the impact of neighborhoods in the presence of inference rules should be studied.

Acknowledgments. This work was supported by grants ANR-22-PECY-0002 (IPoP project), ANR-23-CMAS-0019 (CyberINSA project), and ANR-23-CE23-0032 (DIF-PRIPOS project) funded by the ANR.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bilal Abu-Salih, Muhammad Al-Qurishi, Mohammed Alweshah, Mohammad Al-Smadi, Reem Alfayez, and Heba Saadeh. Healthcare knowledge graph construction: A systematic review of the state-of-the-art, open issues, and opportunities. *Journal of Big Data*, 10(1):81, 2023.
2. Latanya Sweeney. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, 10(5):557–570, 2002.
3. Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.
4. Daniel Kifer and Ashwin Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems (TODS)*, 39(1):1–36, 2014.
5. Aleksa Vukotic, Nicki Watt, Tareq Abedrabbo, Dominic Fox, and Jonas Partner. *Neo4j in action*. Manning Publications Co., 2014.
6. Dan Brickley and R. V. Guha. Rdf schema 1.1. World Wide Web Consortium (W3C) Recommendation, February 2014. W3C Recommendation 25 February 2014.
7. Ian Horrocks, Peter F Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean, et al. Swrl: A semantic web rule language combining owl and ruleml. *W3C Member submission*, 21(79):1–31, 2004.
8. Robert DC Shearer, Boris Motik, and Ian Horrocks. Hermit: A highly-efficient owl reasoner. In *Owled*, volume 432, page 91, 2008.
9. Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography (TCC)*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284, 2006.
10. Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*. Foundations and Trends in Theoretical Computer Science. Now Publishers, 2014.
11. Michael Hay, Chao Li, Gerome Miklau, and David Jensen. Accurate estimation of the degree distribution of private networks. In *2009 Ninth IEEE International Conference on Data Mining*, pages 169–178, 2009.

12. Wei-Yen Day, Ninghui Li, and Min Lyu. Publishing graph degree distribution with node differential privacy. SIGMOD '16, page 123–138, New York, NY, USA, 2016. Association for Computing Machinery.
13. Joseph P. Near and Xi He. Differential privacy for databases. *Found. Trends Databases*, 11(2):109–225, 2021.
14. Borja Balle, Giovanni Cherubin, and Jamie Hayes. Reconstructing training data with informed adversaries. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1138–1156. IEEE, 2022.
15. Jaewoo Lee and Chris Clifton. How much is enough? choosing ε for differential privacy. In *Information Security: 14th International Conference, ISC 2011, Xi'an, China, October 26-29, 2011. Proceedings 14*, pages 325–340. Springer, 2011.
16. Hassan S Al Khatib, Subash Neupane, Harish Kumar Manchukonda, Noorbakhsh Amiri Golilarz, Sudip Mittal, Amin Amirlatifi, and Shahram Rahimi. Patient-centric knowledge graphs: a survey of current methods, challenges, and applications. *Frontiers in Artificial Intelligence*, 7:1388479, 2024.
17. Yuke Hu, Wei Liang, Ruofan Wu, Kai Xiao, Weiqiang Wang, Xiaochen Li, Jinfei Liu, and Zhan Qin. Quantifying and defending against privacy threats on federated knowledge graph embedding. In *Proceedings of the ACM Web Conference 2023*, pages 2306–2317, 2023.
18. Jenni Reuben. Towards a differential privacy theory for edge-labeled directed graphs. *SICHERHEIT 2018*, 2018.
19. Sara Taki, Cédric Eichler, and Benjamin Nguyen. It's too noisy in here: Using projection to improve differential privacy on rdf graphs. In Silvia Chiusano, Tania Cerquitelli, Robert Wrembel, Kjetil Nørvåg, Barbara Catania, Genoveva Vargas-Solar, and Ester Zumpano, editors, *New Trends in Database and Information Systems*, pages 212–221, Cham, 2022. Springer International Publishing.
20. Xiaolin Han, Daniele Dell'Aglio, Tobias Grubenmann, Reynold Cheng, and Abraham Bernstein. A framework for differentially-private knowledge graph embeddings. *Journal of Web Semantics*, 72:100696, 2022.
21. Milad Nasr, Shuang Song, Abhradeep Thakurta, Nicolas Papernot, and Nicholas Carlini. Adversary instantiation: Lower bounds for differentially private machine learning. In *SP*, pages 866–882. IEEE, 2021.
22. Jordi Soria-Comas, Josep Domingo-Ferrer, David Sánchez, and David Megías. Individual differential privacy: A utility-preserving formulation of differential privacy guarantees. *IEEE Transactions on Information Forensics and Security*, 12(6):1418–1429, 2017.
23. Christine M O'Keefe and Anne-Sophie Charest. Bootstrap differential privacy. *Trans. Data Priv.*, 12(1):1–28, 2019.
24. Prottay Protivash, John Durrell, Zeyu Ding, Danfeng Zhang, and Daniel Kifer. Reconstruction attacks on aggressive relaxations of differential privacy. *arXiv preprint arXiv:2209.03905*, 2022.
25. Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Marco Stronati. Geo-indistinguishability: A principled approach to location privacy. In *Distributed Computing and Internet Technology: 11th International Conference, ICDCIT 2015, Bhubaneswar, India, February 5-8, 2015. Proceedings 11*, pages 49–72. Springer, 2015.
26. Konstantinos Chatzikokolakis, Miguel E Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi. Broadening the scope of differential privacy using metrics. In *international symposium on privacy enhancing technologies symposium*, pages 82–102. Springer, 2013.

27. David Liben-Nowell and Jon Kleinberg. The link prediction problem for social networks. In *Proceedings of the twelfth international conference on Information and knowledge management*, pages 556–559, 2003.
28. Bin Yang, Issei Sato, and Hiroshi Nakagawa. Bayesian differential privacy on correlated data. In *Proceedings of the 2015 ACM SIGMOD international conference on Management of Data*, pages 747–762, 2015.
29. Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science, (FOCS)*, pages 94–103. IEEE Computer Society, 2007.